Where to Afforest?

Single and Multiple Criteria Evaluation Methods for Spatio-Temporal Decision Support, with Application to Afforestation

René ESTRELLA

Examination committee:
Prof. dr. ir. D. Berckmans, chair
Prof. dr. ir. J. Van Orshoven, supervisor
Prof. dr. ir. D. Cattrysse, co-supervisor
Prof. dr. D. De Schreye (KU Leuven)
Prof. dr. A. Van Rompaey (KU Leuven)
Prof. dr. ir. B. Muys (KU Leuven)
Prof. dr. C. Romero (UPM, Spain)

Dissertation presented in partial fulfillment of the requirements for the degree of PhD in Bioscience Engineering

November 2015

Doctoraatsproefschrift nr. 1303 aan de Faculteit Bio-ingenieurswetenschappen van de KU Leuven

Uitgegeven in eigen beheer, René Estrella, Celestijnenlaan 200E box 2411, B-3001 Leuven (Belgium)

Acknowledgments

Carrying out a PhD research project is never the work of only one person and I think that such statement is particularly true in my case. At this point I have the pleasant opportunity to recognize all the people that in one way or another contributed to my PhD.

This PhD work was funded by a FLOF scholarship from the Department of Earth and Environmental Sciences, KU Leuven. I want to express my gratitude to the KU Leuven as a whole and, even more in general, to Belgium, for having given me shelter and for becoming my home far from home for the past five years of my life.

My gratitude goes also to all members of the Examination Committee of my PhD. Thank you for taking the time and effort to evaluate my PhD and to give me all those useful suggestions that I am convinced greatly contributed to the improvement of this manuscript.

Thanks Prof. Daniel Berckmans for chairing the preliminary defense of my PhD in such a kind and friendly way. We had personal contact only then, but for me that was enough to realize that you have a lot of experience and skills on coordinating that kind of discussions.

Thanks Prof. Bart Muys. We first met each other in Cuenca, several years before I started my PhD. At that moment it was already very clear to me that you are an extremely kind person, which is something I have had the opportunity to confirm during my multiple stays in Belgium.

Thanks Prof. Carlos Romero, in the first place, for having me during my stay in Madrid in April-May 2012. I am very grateful for your constant willingness to help me and for the expert insights you provided me with while I was there. Your expertise on Multi-Criteria Decision Making methods was immediately evident to me. That is why a few years later I had the fortunate idea of suggesting your name as one of the External Members of the Examination Committee of my PhD. I want to also thank you very much for guiding me in my first steps on Multi-Criteria Decision Methods in general, and on Goal Programming and Compromise Programming in particular, which would result later in a significant phase of my PhD, reflected in Chapters 5 and 6 of this manuscript.

Thanks Prof. Anton Van Rompaey for all your contributions throughout my PhD, especially during our Supervisory Committee meetings. I want to thank you also for always being willing to kindly help me.

Thanks Prof. Danny De Schreye, not only for your contributions to my PhD, but also for your supervision during the development of my MSc. thesis in the Artificial Intelligence program. You always have had kind words of encouragement for me, specially when I was going through challenging circumstances, like giving the oral exam of Fundamentals of Artificial Intelligence or having our intermediate Supervisory Committee meetings. From the first of those occasions I realized that you really understand how a student is feeling and how to help her or him to overcome the challenge.

I have few but special words of gratitude to the co-supervisor of my PhD, Prof. Dirk Cattrysse. I am so grateful to you Professor for all the hard work you have put on my PhD, especially during all our monthly meetings, which proved to be so stressful but at the same time very useful to me. Not to mention the work that implies reading my reports and paper drafts, and always making insightful comments on them. I really admire your intellectual skills. You would always come up with new perspectives on my results that would have hardly occur to me but they always proved to be bright ways of approaching the tasks at hand. I also want to thank you very much for having given me several opportunities when taking your course on Deterministic Decision Models. Thanks for your kind understanding.

I want to express my gratitude to Ward Delabastita, whose work during his MSc. thesis served as the basis for my first publication as a main author, which in turn was the base for Chapter 4 of this manuscript. Besides your thesis work Ward, you were always willing to help me to clarify some aspects related to your results, something I am very grateful for. I want to also thank Liesbeth De Keukelaere and Kaixi Zhang, for whom I had the honor to supervise their MSc. thesis work. Supervising a MSc. thesis proved to be a really enriching experience for me, for which I am grateful. Liesbeth and Kaixi, I hope to have been at the level of the circumstances, and I hope to have managed to contribute in some way to your thesis work. Your work certainly and greatly contributed to my PhD, by providing me with the necessary datasets for Chapter 2 (Liesbeth) and providing the base text, results and analysis for Chapter 3 (Kaixi).

I want to thank all my colleagues and friends at the Division of Forest, Nature and Landscape. Thanks for accompanying me for the past four years of my life. A special thank you goes to Sam. Thanks Sam for translating the abstract of this manuscript to Nederlands, for stubbornly organizing the football games and, most importantly, thanks for the friendship you prized me with.

I also want to thank all my friends from Cuenca who have been accompanying me during the past five years in Leuven. I am very grateful for all the friendly meetings we shared, the good experiences we lived together, and the so many memories that I will keep with me.

A special acknowledgement goes to Pablo Vanegas. Thanks Pablo for having always been encouraging me to undertake postgraduate studies. Without that encouragement I doubt I would have ever started pursuing those goals. Thanks as well for all your support during my MSc. and PhD. studies.

I express my gratitude to Olga Zalamea. Olga, thanks for your company during the past one year and a half. It is amazing that in such a short time we have become close friends. Thanks for your generosity and for coping with my harsh style of showing appreciation to people.

I want to take this opportunity to also thank Jaime Veintimilla. Jaime, thanks for being by my side for the past twenty five years of my life. We first met each other when we were two children trying to find our way in life, and we have been friends ever since. I think that our friendship now has proven that it can endure distance, both in terms of time and space. You are the brother I chose.

I have very special words of gratitude to the supervisor of my PhD, Prof. Jos Van Orshoven. Professor, I decided to write these words between the lines dedicated to my best friends and my family because that is the place I believe you have in my life. At the first time we met each other, almost exactly one decade ago, I immediately notice your kindness and good spirit. I have made nothing but constantly confirm those first impressions ever since. Probably from a Belgian perspective it is always preferable to discriminate well between personal and work relationships, but from my Ecuadorian perspective I regard you as someone much closer to me than a supervisor. There are simply too many things I have to thank you for, but given the limited space I am forced to choose a tiny sample. Thanks for your very kind, supportive and helpful supervision, not only during my PhD, but also during what we used to call in Cuenca “the VLIR6 project”, and also during my MSc. thesis work. Thanks for giving me the great and challenging opportunity of working on a PhD and, in that way, for helping me to set a clear and concrete goal for the past four years of my life. Now it is time for me to look for new goals. I am confident that using what I have learned and following the example of people like you I will manage to find them. A big thanks for all your contributions to my PhD work.

They are too many to even try to list them here, but I firmly believe that you invested much more work on my PhD than what a supervisor is supposed to. Special thanks for your limitless understanding, and for not having given up on me, although I know I was a particularly difficult PhD student to supervise. Thanks for the constant encouragement, for always being willing to listen to me. Thank you for always keeping the door of your office open and for always being available to answer my questions. I simply do not have words to thank you as I should, what I can certainly affirm is that a silent gratitude to you will reside forever in my hearth.

I want to also state my deepest gratitude to my sister Raquel. Raquel, thank you very much for not forgetting me. I do not know whether I deserve it or not, but I have always known that your love has been with me all the time. Since the time we were just a couple of children either playing or fighting on the backyard of our parents' house, I have always felt specially linked to you. Now I know that this link will never be broken by any event or circumstance. I now know that it will endure until the last day of our lives.

I have to switch to Spanish for the last lines of this text, because that is the language in which my hearth can best express itself. Reservé estas líneas para la persona en quien reside el mayor de mis afectos, mi madre. Mamá, su presencia en mi vida ha sido siempre una enorme bendición. Ud. ha sido un ejemplo de fortaleza tan grande que yo no aspiro tan siquiera a acercarme al nivel de fuerza que Ud. ha tenido que demostrar para superar tantos y tan fuertes retos que la vida ha colocado en su camino. Tal vez el más grande de ellos, el hecho de que nos hayamos tenido que separar físicamente de quien fue su esposo y mi padre, y de quien sé con certeza que ha estado a mi lado, hombro con hombro, en todos y cada uno de los retos que conllevan una maestría y un doctorado en el extranjero. Estoy muy agradecido con la vida por el hecho de que Ud. mamá, una vez más dándonos a todos sus hijos un ejemplo de infinita fortaleza, haya sido capaz de resistir el golpe tan duro que supuso la pérdida de la presencia física de su esposo y mi padre. Mamá, espero que la vida me premie con muchos años más de su tan preciada compañía. Ahora recuerdo aquel lejano instante, acerca del cual Ud. me ha contado, en que hace más de 40 años que Ud. y papá decidieron dejar atrás tantas cosas queridas en su amado Gúlag para empezar a enfrentar todos los retos que conllevan una nueva vida en Gualaceo, con el solo propósito de darles más oportunidades de desarrollo a sus hijos. También ahora recuerdo que Ud. me contó que el Jorge lloraba cuando estaba por terminar la escuela en Gualaceo, por el sueño que él tenía, siendo apenas un niño de doce años, de ir a estudiar a Cuenca. Ud. y papá entonces tomaron nuevamente el reto tan duro de dejar atrás una vida y empezar otra en Cuenca. Ahora más que nunca tengo presentes en mi mente las imágenes basadas en sus relatos de las ocasiones en que, hace alrededor de 35 años, sin conocer Cuenca y conmigo

en sus brazos, Ud. subía a sitios altos de la ciudad para tratar de tener una vista panorámica y encontrar así una escuela para sus hijos. Lo único que he hecho yo, desde aquel lejano día en que empecé la escuela hasta ahora que mi doctorado se acerca a su culminación, es tratar de seguir a mi manera el camino que Ud., papá y el Jorge trazaron. Gracias por ser mis ejemplos de vida.

René Estrella
Leuven, November 2015

Abstract

Land evaluation and land use planning are scientific disciplines and professional practices that aim at promoting sustainable territorial development across the globe. One of the cornerstone elements in these disciplines and practices is the notion of land unit. A land unit is an area of land for which the within-unit variability of its functional qualities is smaller than the between-unit variability. With this concept in place, the basic purpose of land evaluation is to establish a matching between the qualities of a land unit and the requirements of a land use type with a view to transferring this information to the subsequent land use planning. This matching can consist in determining the most suitable land units for a given land use type or finding the optimal land use type that should be applied to a given land unit, with the ultimate goal of achieving a high performing, flexible and sustainable land use and management regime for the benefit of society.

Several methods originating from decision sciences have recently been proposed to match land unit's qualities to land use type's requirements. Essentially, decision analysis is about selecting the best option among a number of feasible alternatives. Such selection involves in most cases the consideration of several conflicting objectives. In these situations the application of multi-criteria decision making methods, which perform a trade-off among the considered criteria to find the best alternative, becomes a requirement.

In this dissertation several multi-criteria decision making methods are studied and applied to the spatial planning of new Pinus and Eucalyptus plantations in the southern Andes region of Ecuador. In the context of these studies, the considered criteria are termed land performance attributes, which can be seen as indicators of productive or regulatory ecosystem services. These attributes are categorized as either on-site, when the attribute level is conditioned only by inherent characteristics of the considered land unit (e.g., the stock of carbon stored in the soil), or off-site, when the attribute level depends on the state and behaviour of neighbouring or even distant land units (e.g., the amount of sediment trapped in a land unit depends on the sediment produced, transported and deposited in the land units upstream). A 'per land unit' land evaluation focused on on-site attributes considers a land unit as an independent entity, and assesses its performance without considering any relationship with other land units. The performance of the study region is then merely the sum of the per land unit performances. On the other hand, land performance can be evaluated at a regional scale. In this approach land units are considered to be interacting components of a larger system, e.g., a river catchment. This interaction must be taken into account in the assessment of regional land performance.

In the first part of this study, afforestation planning was tackled using the existing a method called CAMF, which aims at locating the cells within a rasterized river catchment that should be afforested in order to minimize a single, off-site criterion, i.e., the sediment yield of the catchment. We proposed a first variant of this method with the goal of enhancing its efficiency in terms of execution time by approaching the problem from an on-site land evaluation perspective. It was observed that this variant was able to select areas for afforestation in a small fraction of the time required by the original method, while producing almost identical results. In the second variant we replaced the single flow direction algorithm used by CAMF to simulate sediment flow over the study catchment by a multiple flow direction model, in an attempt to make the sediment flow component more realistic. It was shown that after calibration, this new variant produces similar results when compared to WaTEM/SEDEM, a well-tested mechanistic sediment delivery simulation model that was used as a reference.

Next, five existing and one newly proposed multi-criteria decision making methods were explored and evaluated in the context of solving the problem of selecting land units for afforestation with a view to optimize several on-site performance attributes. Despite the strong concepts behind the new Iterative Ideal Point Thresholding method, it was found less applicable to this particular case due to its inability to discriminate between closely related alternatives in a reasonable amount of iterations.

Afforestation planning is not necessarily restricted to selecting areas within a region to establish forests. From a more strategic perspective the goal can be to determine where forestry land use types should be established and how they should be managed to maximize their benefits considering the full region as one entity. With a view to provide decision support tools that are useful in such strategic contexts, a Mathematical Programming model was formulated, implemented and its applicability was demonstrated. In a first stage, the model was applied to determine the full set of land unit-land use type combinations (or land use configuration) that should be established in a river catchment in order to optimize regional land performance. Regional land performance was expressed as the integration over the study area of the level achieved by a number of conflicting on-site attributes. An additional requirement set for the solutions produced by this model was that the achievement levels of the considered attributes are as uniform as possible, avoiding solutions that perform well for certain attributes and very poorly for the others.

To enhance the flexibility of the approach described above, in a second stage, the Mathematical Programming model was further elaborated to allow land use changes to occur at fixed time intervals. This enhancement introduces the temporal dimension in the concept of land use configuration and extends it to the notion of land use trajectory configuration. Herewith, a land use trajectory is defined as a particular sequencing of several pre-defined land use types, defining the land use change that may or may not occur at the end of every interval within a considered time span. The result of the Mathematical Model when applied to the studied problem was a land use trajectory configuration that provides information about how land use should evolve throughout a period of time in order to obtain, at the end of this period, an optimal and balanced, temporally cumulated regional land performance.

The application of these two Mathematical Programming models to the Southern Andes study area proved that a combination of techniques borrowed from multi-criteria decision analysis and Operations Research can result in valid methods for determining static or dynamic land use configurations that favour land performance enhancement while keeping a balance among the achievement levels of the considered attributes.

The methods proposed in this dissertation must be seen as new or enhanced tools for land use planning in general and afforestation planning in particular. The aim of these methods or of any decision support tool is not to replace but rather to support the sensible judgement by stakeholders where it regards complex spatio-temporal decision problems.

Samenvatting

Landevaluatie en landgebruiksplanning omvatten wetenschappelijke disciplines en professionele praktijken die overal ter wereld ingezet worden om duurzame ruimtelijke ontwikkeling te promoten. Eén van de sleutelbegrippen in deze disciplines en praktijken is 'landeenheid'. Een landeenheid is een afgebakend deel van het aardoppervlak waarbinnen de functionele eigenschappen minder variabel zijn dan het geval is tussen de eenheden. Het doel van landevaluatie is de informatie te leveren voor een optimale afstemming van de eigenschappen van een landeenheid enerzijds met de vereisten van het beschouwde landgebruikstype anderzijds zodat landgebruiksplanners hiermee aan de slag kunnen. Deze afstemming of 'matching' kan bestaan uit het bepalen van de meest geschikte landeenheden voor een gegeven landgebruikstype, of uit het vinden van het optimale landgebruikstype voor een gegeven landeenheid.

Meerdere methodes die ontwikkeld zijn in het domein van de beslissingswetenschappen zijn recent aan bod gekomen binnen de landevaluatie. Beslissingsanalyse bestaat in essentie uit het selecteren van de beste optie uit een verzameling van meerdere haalbare alternatieven. Deze selectie houdt vaak het in acht nemen van meerdere conflicterende objectieven in. In deze gevallen wordt het toepassen van multi-criteria beslissingsondersteunende methodes, die een afweging maken tussen de verschillende criteria om het beste alternatief te vinden, een vereiste.

In deze verhandeling worden verschillende multi-criteria beslissingsondersteunende methodes bestudeerd en toegepast op de ruimtelijke planning van nieuwe Pinus- en Eucalyptus-aanplantingen in de zuidelijke Andes-regio van Ecuador. In de context van deze studie komen de beschouwde criteria overeen met performantie attributen, die ook gezien kunnen worden als indicatoren van voorzienende of regulerende ecosysteemdiensten. Deze attributen kunnen gecategoriseerd worden als ofwel on-site, wanneer het attributenniveau enkel door lokale factoren van een landeenheid bepaald wordt (bv. de koolstofvoorraad in de bodem), ofwel off-site, wanneer het attributenniveau afhankelijk is van

de toestand en het gedrag van naburige of zelfs verafgelegen landeenheden (bv. de hoeveelheid sediment die wordt vastgehouden in een landeenheid, die afhankelijk is van het sediment dat bovenstrooms geproduceerd, getransporteerd en afgezet wordt). Zowel on- als off-site attributen kunnen beoordeeld worden op twee verschillende manieren. Een 'per landeenheid' evaluatie werkt met on-site attributen, beschouwt een landeenheid als een onafhankelijke entiteit en stelt zijn performantie vast zonder de relaties met andere landeenheden in rekening te brengen. De performantie van een studiegebied is dan louter de optelsom van de per-landeenheid performanties. Daarnaast kan performantie geëvalueerd worden op regionale schaal. Hierbij worden landeenheden beschouwd als interagerende componenten van een groter systeem, i.e. het studiegebied (bv. een rivierbekken). Deze interacties moeten in rekening gebracht worden bij de beoordeling van de regionale performantie.

In het eerste deel van deze studie stond de bestaande methode CAMF centraal. CAMF laat toe cellen te selecteren in een gerasterd rivierbekken, die meest in aanmerking komen voor bebossing op basis van de minimalisatie van één enkel, off-site criterium, met name het sedimentverlies van het bekken. Een eerste variant van deze methode werd voorgesteld met het oog op het verhogen van de reken-efficiëntie, door de opdracht te herformuleren als een on-site landevaluatieprobleem. Er werd vastgesteld dat deze variant toeliet om gebieden voor bebossing te selecteren in slechts een fractie van de tijd die de originele methode nodig had, terwijl de resultaten haast identiek waren. Voor de tweede variant vervingen we het enkelvoudige door een meervoudige stromingsrichtingsalgoritme, in een poging om de sedimentafvoer realistischer te modelleren. Deze nieuwe variant leverde gelijkaardige resultaten als een uitgebreid geteste, mechanistische methode (WaTEM/SEDEM) die als referentie gebruikt werd.

In het volgende deel werden vijf bestaande en één nieuw ontwikkelde methode voor multi-criteria analyse verkend en geëvalueerd met het oog op het kiezen van locaties voor bebossing waarbij meerdere on-site landperformantie attributen geoptimaliseerd werden. Ondanks de conceptuele sterkten van de nieuwe Iterative Ideal Point Thresholding methode, werd gevonden dat deze methode minder toepasbaar is omwille van het hoge aantal iteraties dat nodig is om sterk gelijkende alternatieven te onderscheiden.

Bebossingplanning is niet noodzakelijk beperkt tot de selectie van te bebossen gebieden binnen een grotere zone. Vanuit een meer strategisch perspectief kan het doel zijn om te bepalen waar bos geplant dient te worden en hoe het te beheren om de performantie te optimaliseren voor het volledige beschouwde gebied. Met het oog op het aanleveren van beslissingsondersteunende instrumenten die bruikbaar zijn in zulke strategische contexten werd een mathematisch programmeermodel geformuleerd, geïmplementeerd en toegepast.

In een eerste fase werd dit model gebruikt om de volledige verzameling van landeenheid-landgebruikstype combinaties (of landgebruiksconfiguratie) te bepalen die leidt tot optimale performantie op de regionale schaal. De regionale landperformantie werd uitgedrukt als de integratie over het hele studiegebied van het niveau dat bereikt wordt door meerdere conflicterende on-site attributen. Een bijkomende vereiste die gesteld werd aan de oplossingen die voortkomen uit dit model was dat de behaalde niveaus van de beschouwde genormaliseerde attributen zo uniform als mogelijk zijn, om zo configuraties te vermijden die goed presteren voor bepaalde attributen maar zwak voor de overige.

Om de flexibiliteit van de hierboven beschreven methode te verbeteren werd, in een tweede fase, het mathematisch programmeermodel verder uitgewerkt om toe te laten dat landgebruiksveranderingen optreden na vaste tijdsintervallen. Deze verbetering introduceert de temporele dimensie in het concept landgebruiksconfiguratie en breidt het uit naar deze van landgebruikstrajectconfiguratie. Hierbij wordt een landgebruikstraject gedefinieerd als een sequentie van meerdere, vooraf gedefinieerde landgebruikstypes en geeft dit aan welke landgebruiksveranderingen worden aangeraden aan het eind van ieder interval binnen de beschouwde tijdshorizont. Het resultaat was een landgebruikstrajectconfiguratie die informatie geeft over hoe het landgebruik dient te evolueren doorheen een periode om, aan het eind van deze periode, een optimale, evenwichtige, in de tijd gecumuleerde regionale landperformantie te bekomen.

De toepassing van deze twee mathematische programmeermodellen op het studiegebied in het zuiden van de Ecuadoraanse Andes-regio toonde aan dat een combinatie van technieken geleend uit multi-criteria beslissingsanalyse enerzijds en operationeel onderzoek anderzijds kan leiden tot bruikbare methodes om statische en dynamische landgebruiksconfiguraties te bepalen die de algehele landperformantie verbeteren terwijl een evenwichtig prestatieniveau van de onderliggende performantie-attributen gegarandeerd is.

De methodes die in deze verhandeling bestudeerd en verder ontwikkeld werden dienen gezien te worden als nieuwe of verbeterde instrumenten voor landgebruiksplanning in het algemeen en bebossingsplanning in het bijzonder. Het doel van deze methoden en –bij uitbreiding- van ieder beslissingsondersteunend instrument, is niet om het zinnige oordeel van beslissingsnemers te vervangen maar eerder om hen ondersteuning te bieden, vooral bij complexe tijdruimtelijke beslissingsproblemen.

List of Abbreviations

AHP	Analytic Hierarchy Process
asl	above sea level
BOC	Biomass Organic Carbon
CAMF	Cellular Automata based method for Minimizing Flow
DEM	Digital Elevation Model
DEMON	Digital Elevation Model Networks
ELECTRE	ELimination Et Choix Traduisant la REalité
FAO	Food and Agriculture Organization
FD8	Fractional Deterministic 8
IIPT	Iterative Ideal Point Thresholding
iLUT	initial Land Use Type
IP	Integer Programming
LP	Linear Programming
LUT	Land Use Type
MCDM	Multi-Criteria Decision Making
MFD	Multiple Flow Direction
PROMETHEE	Preference Ranking Organisation METHod for Enrichment Evaluations
RRMSD	Relative Root Mean Square Deviation
RUSLE	Revised Universal Soil Loss Equation
SFD	Single Flow Direction
SMAA	Stochastic Multi-criteria Acceptability Analysis
SOC	Soil Organic Carbon
USD	United States Dollars

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem statement

The recognition that territorial developments are often unsustainable has given rise to a variety of disciplines aimed at alleviating the negative effects of uncontrolled or injudicious land cover and land use changes. One of such disciplines is land use planning, which in simple terms can be defined as the allocation of the appropriate land use types (LUTs) to area units within a region. A closely related area of study is land evaluation. The principles of land evaluation were introduced in the 1970s (FAO (1976)) and updated in the 2000s (FAO (2007)). Both land use planning and land evaluation techniques start by stratifying the region of interest into land units. Land units are portions of land for which the within-unit variability regarding characteristics or qualities is smaller than the between-unit variability. The assignment of a specific LUT to a land unit is performed by matching the characteristics or qualities of the land unit to the requirements of the candidate LUT. This matching is performed with a view to achieve the primary goal of land evaluation, which is to manage land in an improved and sustainable way for the benefit of people (FAO (2007)).

One of the LUTs that are frequently considered as a viable option for sustainable development especially in rural areas is forestry. Afforestation is a particularly appealing alternative for degraded land, where agriculture is no longer feasible or economically profitable. With regard to soil erosion and sediment issues, forests are known to protect the soil surface from the direct impact of rain drops, reducing their potential to detach soil particles. Afforestation is also known as a measure to decrease the ability of runoff to transport sediment (Morgan (2009)).

Any afforestation must be carefully planned in order to achieve the desired land performance. For instance, land units can be ranked according to their suitability for afforestation with a given tree species. The most suitable land units can be selected from this ranking as the optimal sites for afforestation. This approach can be used to answer the question of where a forest should be established. From this perspective afforestation planning becomes an instance of a site location problem (e.g., Vanegas et al. (2008)). On the other hand, different tree species can also be ranked according to the extent at which its biophysical requirements are fulfilled by a given land unit. The highest ranked tree species can then be chosen to afforest the land unit at stake.

Another fundamental factor in any afforestation planning project is the temporal dimension. Given that changes in land performance under forest are related to tree growth and are, hence, gradual, the objectives of afforestation planning cannot be confined to the immediate future. Therefore, it is no longer enough to determine a single tree species to afforest a given land unit. It becomes necessary to determine the most appropriate sequence of tree species that should be used to afforest the land unit at stake during a certain time period. The same reasoning can be applied for any other LUTs. In this context, time can be considered in absolute terms, i.e., a period between specific years (e.g., 2015-2045), or it can be defined relatively to a reference point (year 0). Considering the temporal dimension in absolute terms brings along additional challenges, such as the need of determining the expected performance of land under a future, uncertain climate.

In addition to the traditional land use planning and evaluation techniques initially introduced in FAO (1976) and FAO (2007), more recent approaches have been proposed either to locate suitable sites for a given LUT or to determine the most appropriate LUT for a given site according to its expected performance. Most of these novel techniques address the questions stated above as decision problems. According to Romero and Rehman (2003) solving a decision problem involves the selection, performed by the decision maker, of the best option from the set of feasible alternatives. The selection of the best alternative may or may not be preceded by a ranking of all alternatives. The set of decision alternatives can be finite or infinite. From the definition of decision problem, the selection of the best alternative among the feasible ones implicitly involves the application of optimization techniques.

The definition of the best alternative depends on the objectives or criteria that are relevant to the decision maker. If the decision maker uses a single criterion to define the optimal alternative, then a simple ranking procedure, according to the considered criterion, would suffice to find the best alternative when the set of alternatives is finite. When only one criterion is considered and the set of alternatives is infinite, techniques taken from Mathematical Programming

(Winston and Goldberg (2004)) like the Simplex algorithm (proposed by George Dantzig in 1947), can be used to find the optimal alternative.

Different though related techniques are required when multiple conflicting criteria are at stake. The term conflict is used in this context to refer to the situation in which improving the level of a certain criterion decreases the level of one or more of the other criteria. Such a problem is an instance of a multi-criteria decision problem. Every decision alternative can be considered as a vector in a multi-dimensional space, where each dimension corresponds to a separate criterion. From this perspective every coordinate value of an alternative corresponds to the level achieved by that alternative for one of the considered criteria. For any decision problem to qualify as multi-criteria it is required that conflict exists between any pair of criteria. When such conflict among criteria is note present, then there exists at least one decision alternative for which all its criteria levels achieve the absolute optimum. These alternatives can be immediately selected as the optimal one using simple queries. Multi-criteria decision techniques are not required in this case. When conflict among criteria does exist, on the other hand, the application of Multi-Criteria Decision Making (MCDM) methods becomes a necessity.

MCDM methods are techniques devised to find optimal feasible solutions while trading off several conflicting criteria. Since in a multi-criteria decision problem there are no alternatives for which all criteria reach the absolute optimum, a compromise must be made to find those alternatives that are as close to the optimum as possible. A multitude of MCDM methods have been proposed in the last decades. Some of these methods are targeted to solve problems with a finite number of decision alternatives, like the Analytic Hierarchy Process (AHP, Saaty (1977)), ELECTRE (ELimination Et Choix Traduisant la REalité, Roy (1991)), PROMETHEE (Preference Ranking Organisation METHod for Enrichment Evaluations, Brans and Vincke (1985)) and IIPT (Gilliams et al. (2005b); De Meyer et al. (2013); Estrella et al. (2014b)). On the other hand, the two most prominent MCDM methods aimed at solving decision problems with an infinite number of alternatives are Compromise Programming (Zeleny (1973); Yu (1973)) and Goal Programming (Charnes et al. (1955); Charnes and Cooper (1957)). MCDM methods for problems with a finite alternative set allow to rank the (typically small) number of alternatives, while the application of a MCDM method devised for problems with an infinite number of alternatives results in a Mathematical Programming model that can be solved using techniques borrowed from Operations Research (Romero and Rehman (2003)).

In this dissertation the criteria considered when applying the different decision making methods are related to ecosystem services, i.e., the goods and services that humans experience from land based ecosystems (MEA (2005)). Such ecosystem services are closely related to land performance, expressed as measurable attributes of land units. Soil erosion and sediment transport and delivery are examples of performance attributes, which will determine the ecosystem services provided by a territory, e.g., soil productivity for agriculture.

Performance attributes are said to be on-site if the level reached by them depends only on local factors pertaining to a land unit. An example of an on-site land performance attribute is the stock of soil organic carbon (SOC) present in a given land unit at a specific point in time. On the other hand, when the performance attribute level is determined by (changes in) the state of neighbouring or even distant areas, then such attribute is considered to be off-site. For instance, the amount of sediment delivered to a river depends on processes happening in the whole upstream area, therefore sediment delivery is an instance of an off-site performance attribute. The mutual influence between land units that is inherent to the definition of any off-site attribute is known as spatial interaction. For example, modelling sediment flow within a river catchment relies heavily in the notion of spatial interaction, since the amount of sediment present in a land unit at any point in time depends on the erosion produced and the amount of sediment transported and deposited in upstream units.

The problem of determining which LUTs to apply in a region in order to optimize its land performance can be tackled at different scales. Local or 'per land unit' optimization takes place when the multi-dimensional performance of land units is optimized independently from each other. When devising an optimal land use plan for certain region, the application of a 'per land unit' optimization boils down to selecting the LUT that would result in an optimal performance for each individual land unit. Land performance optimization integrated at a regional scale, on the other hand, aims at achieving the best performance of the full area as a whole. From this perspective land units are considered to be interacting components of a system, rather than independent elements. This means that changes in the state of a given spatial unit has an impact on the regional performance, which in turn will affect decisions made regarding the other land units.

The question of which LUTs to apply in a region in order to optimize on-site performance attributes can be tackled either on a 'per land unit' basis or at a regional scale. 'Per land unit' optimization of on-site performance attributes would amount in this case to select, for each separate land unit, the LUT that would result in the optimum multi-dimensional performance. A regional performance optimization, on the other hand, would involve integrating the on-site performance levels of individual land units over the whole study area. In this case, a regional performance level corresponds to an aggregate measure that expresses the performance of the full study region as a whole. From this perspective land units are considered to be interacting components of a system, rather than independent elements. This means that changes in the state of a given spatial unit has an impact on the regional performance, which in turn will affect decisions made regarding the other land units.

Optimizing off-site attributes can also be performed either in a 'per land unit' fashion or at a regional scale. The problem of optimizing the amount of sediment present in every individual land unit at a specific point in time is an example of 'per land unit' optimization of an off-site performance attribute, since the amount of sediment in a land unit depends on factors related to other land units and every land unit is considered separately from each other. The problem of minimizing the total sediment delivered to the river is an example of optimization of off-site attributes at regional scale, since in these cases the total performance of the full region is expressed as a single aggregate value.

In all the problems sketched above, whether they involve on- or off-site performance attributes, and 'per land unit' or regional optimization is at stake, the temporal dimension is implicitly present. In all cases decisions are made about where a land use change should take place or which LUT should be established at a reference point in time. Such decisions are determined by the impact that the applied land use changes will have on the expected land performance at a future point in time. In this sense, all performance attributes considered in this study can be said to be 'off-time'.

Based on the categorizations established above, which refer to decision problems in which the goal is the optimization of land performance attributes and, ultimately, ecosystem services those attributes are linked to, several possible problem instances can be defined. Specifically, problems combining multiple on- and/or off-site attributes and requiring the optimization of land performance 'per land unit' or at a regional scale, can be defined as relevant depending on the interests and goals of the decision maker. Although some methods have been proposed in the past to tackle concrete instances of these problems (e.g., Vanegas (2010)), in general there is a gap regarding tools, methods and techniques to support site location and the design of land use plans that take on-site and off-site performance attributes into account. In other words, the complete set of possible problems derived from the combinations above is far from having been fully addressed. This dissertation is meant to be a step forward in covering this so-defined set of decision problems involving land use planning for the enhancement of ecosystem services. With this aim in mind, three concrete instances of such afforestation planning problems were addressed throughout this research: (i) optimization of a single off-site performance attribute at a regional scale, (ii) 'per land unit' optimization of multiple on-site performance attributes, and (iii) optimization of multiple on-site performance attributes at a regional scale.

1.2 Research objectives

The general research objective of this dissertation is the formulation, application and evaluation of computationally-efficient decision making methods to support land use planning in general and site location for afforestation in particular. In this respect and in line with the problem stated in Section 1.1 both single and multiple criteria decision problems are addressed, on- as well as off-site performance attributes are considered and attention is given to both the land unit and regional scale. This research does however not tackle the integration of all these methods though ultimately required. The term 'planning support' in this context refers to provide the decision maker with hints or suggestions about the land use configurations, with emphasis on forestry, that would optimize land performance, with a view to alleviate or even reverse the trend of unsustainable development in rural areas.

The first specific objective is targeted to the problem of locating sites within a river catchment that should be afforested with a given tree species in order to optimize a single off-site performance attribute, i.e., the sediment yield of the river catchment. Since the attention here is focused on an aggregate performance measure for the full catchment, i.e., sediment yield, this is an instance of optimization at a regional scale. To tackle this problem, variants of an existing site location method are formulated, implemented and tested, and their performance in terms of accuracy and computational efficiency is assessed.

The second specific objective is focused on locating sites for afforestation with a given tree species in order to optimize land performance expressed in terms of a number of conflicting, on-site criteria. In particular, the goal is to produce a ranking of a set of land units in which the study area is stratified, according to their suitability for the considered tree species. Numerous MCDM methods that can be used to solve this decision problem are available have been proposed in the literature. A subset of such methods are selected and implemented, and their relative performance is tested and analyzed in the context of this site location problem.

The third specific objective aims at determining land use configurations to be applied over the study area in order to optimize land performance both on a 'per land unit' basis and integrated at a regional scale. The term land use configuration refers to a particular distribution of several predefined LUTs over the set of land units within a region. To this purpose, three MCDM methods are either fine-tuned or formulated: on the one hand, the Iterative Ideal Point Thresholding (IIPT) method, which is suited for decision problems with a finite number of alternatives in which 'per land unit' optimization is sufficient, and, on the other hand, two Integer Programming (IP) models derived from the application of Compromise Programming and Composite Programming. The IP models allow to determine the land use configuration that should be established in the study area when regional land performance is at stake. The criteria considered at this stage correspond to several conflicting, on-site performance attributes. Additionally, as a first step towards the inclusion of the temporal dimension in afforestation planning problems, this objective also involves determining the way in which several LUTs should be sequenced at fixed time intervals in each of the land units of a study area in order to attain an optimal land performance at regional scale after a predefined time span.

1.3 Overview of the manuscript

In the current chapter, the general problem is stated and its relevance is emphasized. Additionally, the research objectives are defined.

In Chapters 2 and 3 the problem of locating afforestation sites (cells from a raster dataset) to optimize a single off-site criterion is addressed. In particular, the criterion at stake is the sediment yield of the river catchment under study. Since in these chapters the land performance of the full catchment is aggregated into a single sediment yield value, this is an instance of performance optimization at a regional scale. Both chapters are based on the Cellular Automata based method for the Minimization of Flow (CAMF, Vanegas et al. (2012)).

The first part of Chapter 2 presents an analysis of the computational efficiency of CAMF. At this stage, the CAMF algorithm as originally introduced was implemented and tested. Several performance aspects were measured and analyzed during those tests. In the second part, a simplified CAMF variant that reduces the requirements in terms of execution time while maintaining the accuracy of the results is formulated, tested and assessed with respect to the performance of the original version of this method.

In order to compute the sediment yield of a river catchment CAMF incorporates a sediment flow simulation component, which is based on a Single Flow Direction (SFD) model. Chapter 3 proposes an elaboration to CAMF that consists in shifting the sediment flow simulation component from Single to a Multiple Flow Direction (MFD) model. Such elaboration is expected to improve the way in which CAMF represents real sediment flow processes. After implementing this CAMF variant, its parameters are calibrated using a different sediment transport and delivery model (WaTEM/SEDEM) as a reference. Finally, the accuracy of the MFD variant of CAMF is assessed for several afforestation scenarios with respect to the reference model.

In Chapter 4 a shift is made from single to multiple criteria decision support. This chapter starts presenting a brief overview of the fundamental concepts involved in decision analysis. In a second part, an exploration of available MCDM methods is reported. A sample of six MCDM methods is then chosen and evaluated in the context of a site location problem (land units that should be afforested) involving the optimization of land performance expressed as several conflicting on-site attributes. Since every land unit is evaluated independently from each other, the problem tackled in this chapter is an instance of 'per land unit' optimization. This exploration and evaluation of MCDM methods is the basis for identifying the existing options that are useful to solve the problems formulated in subsequent chapters.

Chapters 5 and 6, like Chapter 4, also focus on the notion of land unit as the basic entity, although the way in which the study region is stratified into land units is different. Similarly to Chapter 4, the problems tackled in Chapters 5 and 6 involve the optimization of multiple conflicting, on-site criteria.

In Chapter 5 two of the methods studied in Chapter 4, namely IIPT and Compromise Programming, are applied to determine the land use configuration that optimizes land performance expressed as a collection of on-site attributes. IIPT is applied to each land unit to determine the LUT that should be applied to optimize the land unit's performance. Compromise Programming is also applied to formulate an IP model that determines optimal land use configurations, that is the ways in which several LUTs should be distributed over the study area in order to optimize regional land performance. The same problem was tackled using a MCDM method closely related to Compromise Programming. This method is called Composite Programming, and it is applied to determine near-to-optimal land use configurations that favours balance regarding the levels achieved for all the considered criteria. The results produced by these three methods are then assessed in relative terms.

Chapter 6 further elaborates on the Composite Programming model formulated in Chapter 5. The fundamental notion in Chapter 6 is land use trajectory. A land use trajectory is defined as a sequence of LUTs that are applied in a land unit during a pre-defined time span. Such a sequence indicates whether a land use change should occur or not at fixed intervals within the time span. Based on these definitions, the Composite Programming model formulated in this chapter is applied to determine the trajectory configuration (concrete distribution of land use trajectories over the study region) that results in an optimal land performance (collection of on-site attributes) at a regional scale after a specified time span. Besides applying this model, its sensitivity to uncertainty in the input data and to systematic variations introduced in its parameter settings is assessed. Moreover, the possibility of incorporating land performance thresholds in the model is demonstrated.

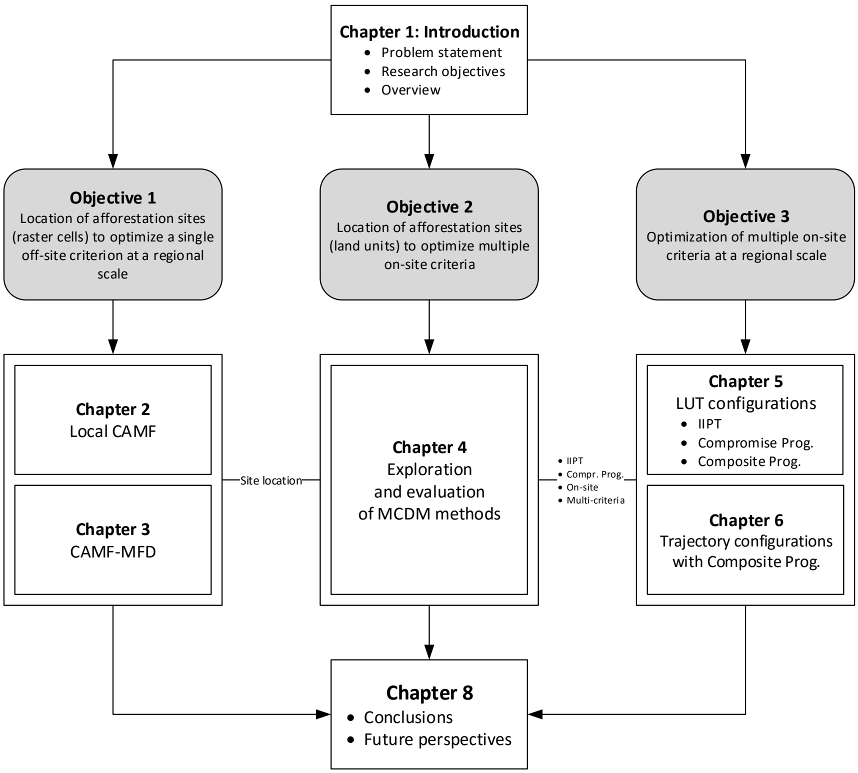

Figure 1.1: Overview of this dissertation.

Finally Chapter 7 contains a general discussion and summarizes the main findings and insights obtained throughout this research. Additionally, it points out a number of facets of this research area as potential items of interest in future research and development.

Figure 1.1 shows a schematic overview of the structure of this manuscript.

Chapter 2

Locating afforestation sites to optimize a single off-site criterion

This chapter is based on:
Estrella, R., Vanegas, P., Cattrysse, D., Van Orshoven, J. (2014). Trading off accuracy and computational efficiency of an afforestation site location method for minimizing sediment yield in a river catchment. In *Proceedings of GEOProcessing 2014: The Sixth International Conference on Advanced Geographic Information Systems, Applications, and Services*, 94-100.

2.1 Introduction

Soil erosion is a common problem in tropical mountainous regions. In such regions, rainfall typically produces high levels of runoff, which in turn causes the soil to be eroded and, as a consequence, large amounts of sediment are produced, transported and deposited (Molina et al. (2008)). This often leads to the undesirable result of degraded soil, i.e., soil with severely limited performance in terms of fertility and productivity. A second major negative consequence caused by soil erosion occurs when the sediment produced is delivered to the river system of a catchment. This sediment will be partially transported so that it will eventually reach the outlet of the catchment. This process is a critical factor when there exists a dammed reservoir downstream the river, since the sediment input to such infrastructures might produce high costs for sediment removal and a shortening of the reservoir lifespan given the resulting loss of capacity (Palmieri et al. (2001)). These factors make the study of sediment flow in mountainous regions crucially important.

One measure that has proven useful to take control over soil erosion and sediment transport is vegetation cover (FAO (1980); Morgan (2009)). According to Yu et al. (2006) vegetation can alleviate soil erosion in several ways. For instance, ground vegetation weakens and diffuses the gravitational energy of runoff, its root system increases the porosity of the soil and interlaces in the soil, enhancing its erosion control capability. In the particular case of forests, besides the benefits mentioned for any kind of vegetation, their canopy can serve as an interceptor of rain drops, weakening their potential for soil particles detachment. Such benefits are enhanced when afforestation is technically planned and based on sufficient scientifically sound information. Typically, when planning an afforestation project, several criteria are to be considered simultaneously. Some of these criteria may pertain to the local performance of areas within the study region. This type of criteria are referred to as on-site. An example of on-site criteria is the amount of carbon sequestered both in soil and in biomass. On the other hand, some criteria can be related to the effect that changes in the state of a given area produce in the state of neighbouring or even distant areas within the study region. These criteria are classified as off-site. Sediment delivery to the river and sediment yield of a river catchment are examples of this type of criteria. Both on-site as well as off-site criteria allow forestry planners to discriminate between suitable and unsuitable alternatives, e.g., selecting sites for afforestation, choosing the species to be planted, or deciding when to harvest the forest.

The term site location for afforestation used throughout this chapter and dissertation refers to determining the exact locations in which trees should be planted. In this specific case, decision alternatives correspond to candidate sites within a river catchment that are available to be afforested. Only areas under agriculture and pasture are considered as candidates for afforestation. The areas covered by other LUTs, like tropical Andean wetlands, natural woody vegetation, or pre-existing forest were excluded from the candidate cells, because it was considered more convenient, for environmental reasons, that these areas remain undisturbed. A single off-site factor, the amount of sediment at the outlet of the catchment, or sediment yield, was chosen as the decision criterion. Since the study regions are represented by raster datasets, the problem amounts to selecting the subset of cells of predefined size that should be afforested in order to minimize the sediment yield of a river catchment.

A computational iterative method to tackle this problem was proposed and tested for small catchments by Vanegas et al. (2012). This method aimed to select, at each iteration, the cell(s) that, in case of being afforested, would produce the maximum reduction in sediment yield. The name Cellular Automata-based method for Minimizing Flow (CAMF) was used to refer to this method. To select a cell or cells at each iteration, CAMF computes the sediment yield reduction that would be produced considering that every candidate cell is afforested separately. This sediment yield reduction values is then used to build a ranking from which the optimal cell(s) is (are) selected.

Some limitations were identified in CAMF during the literature review and preliminary tests. One of these limitations is the fact that scoring cells and building the ranking are relatively expensive procedures in terms of computation time. A second limitation is that there is a high probability that only one cell is selected at each iteration, so that many iterations of CAMF are necessary in order to select the required number of cells. This undesirable combination of repeating many times a computationally expensive procedure might restrict the applicability of CAMF when dealing with high resolution datasets that cover extensive study areas.

The work in this chapter aimed at providing insights about several aspects of CAMF. First, the performance of CAMF was examined as a function of the size of the database to which it is applied and of the number of cells to be selected. This analysis produced indicators about the applicability of CAMF to large databases, which are frequently found in natural resources management projects. This goal was meant to complement the work reported in Vanegas et al. (2012), where only very small, sample databases were used during tests. The second general aim was to propose a variant of CAMF that addresses its limitations to drastically reduce its execution time while preserving the quality of the results it produces.

2.2 Materials and methods

2.2.1 Study regions and datasets

Two study regions were considered in this study, the Paute river catchment and one of its subcatchments, Tabacay. Three raster geodatabases representing these regions were used for testing CAMF and its proposed variant. These geodatabases are stored using the ArcInfo ASCII grid format with a cell resolution of 30x30 m.

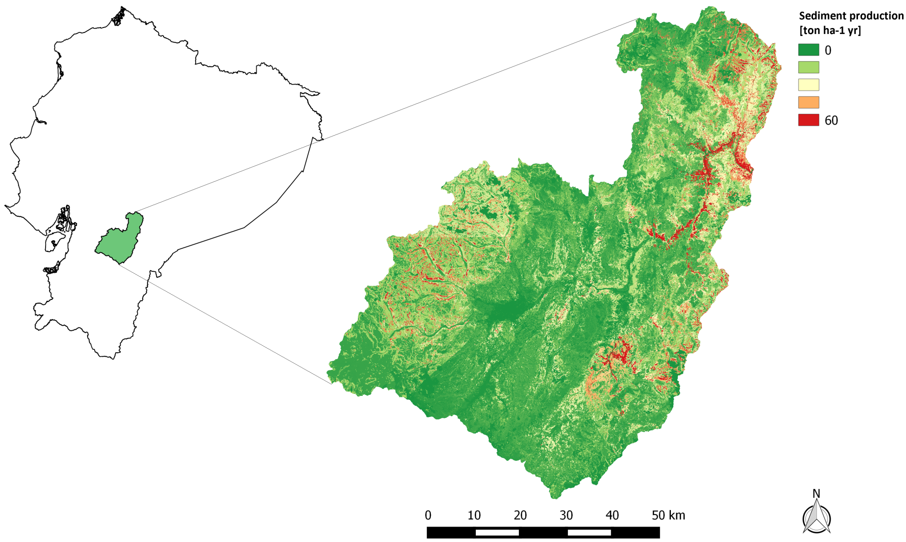

Figure 2.1: Location in Ecuador and sediment production under initial land use (source: De Keukelaere (2013)) of the Paute catchment.

The Paute river catchment

The hydrographical catchment of the river Paute is located in the south of the Andes mountain range in the republic of Ecuador (north-west of the South American continent). Its area is 5055 km^2, with altitudes varying between 1591 and 4651 m asl. The areas under agriculture and pasture in this catchment correspond to a total of 1483 km^2, which represents around 30% of its full area. Several dammed reservoirs that are part of one of the most important Ecuadorian hydroelectric complexes are located within its boundaries.

A sediment production dataset of the Paute catchment retrieved from De Keukelaere (2013) is used in this chapter as input data. This dataset was derived using the equation introduced by Molina et al. (2008), who states that sediment production in catchments in the southern Andes of Ecuador mainly depends on the fractional vegetation cover and on the presence of argillaceous rocks. Figure 2.1 shows this dataset along with the location of the Paute catchment in Ecuador.

The Tabacay river catchment

The hydrographical catchment of the river Tabacay is situated in the province of Cañar (southern Andes of Ecuador). The area of this catchment is 66.52 km^2 and its altitude ranges between 2481 and 3732 m asl. Despite its tropical location its climate is characterized by daily average temperatures between 8 and 19 degrees Celsius, with strong daily fluctuations, which is typical for the Andes region.

The Tabacay catchment can be considered as a representative region for the southern Andes of Ecuador as a whole (Wijffels and Van Orshoven (2009)). The Tabacay river is used as the only source for provision of drinking water to the city of Azogues, the capital of the province of Cañar. Agriculture and pasture cover 24 km^2 (39% of the total catchment area). The Tabacay catchment has a long history of unsustainable land use practices even in highlands with steep slopes. Such practices disturb the soil ecosystems found in those areas. For instance, the widespread agricultural use of the land in those areas leads to large amounts of sediment produced and transported towards the river system, which results in severe land and river degradation (Pimentel et al. (1995)) and reservoir siltation (Palmieri et al. (2001)).

The location of the Tabacay catchment in Paute and its sediment production map, which extracted from the sediment production dataset of Paute, are displayed in Figure 2.2.

The Tabacay500 dataset

An additional, smaller "region" that corresponds to an area of 1.7 km^2 around the outlet of the Tabacay catchment was considered in this chapter. The codename Tabacay500 was chosen for this dataset because it comprises 500 candidate (for afforestation) cells (total number of cells: 1892). The location of Tabacay500 within the Tabacay catchment and its sediment production are displayed in Figure 2.3.

2.2.2 The Cellular Automata based method for Minimizing Flow (CAMF)

Vanegas et al. (2012) proposed the Cellular Automata based method for Minimizing Flow (CAMF) to locate sites within a river catchment that should be afforested in order to minimize sediment delivery to the river system. Spatial interaction among cells in the raster datasets representing the catchment is

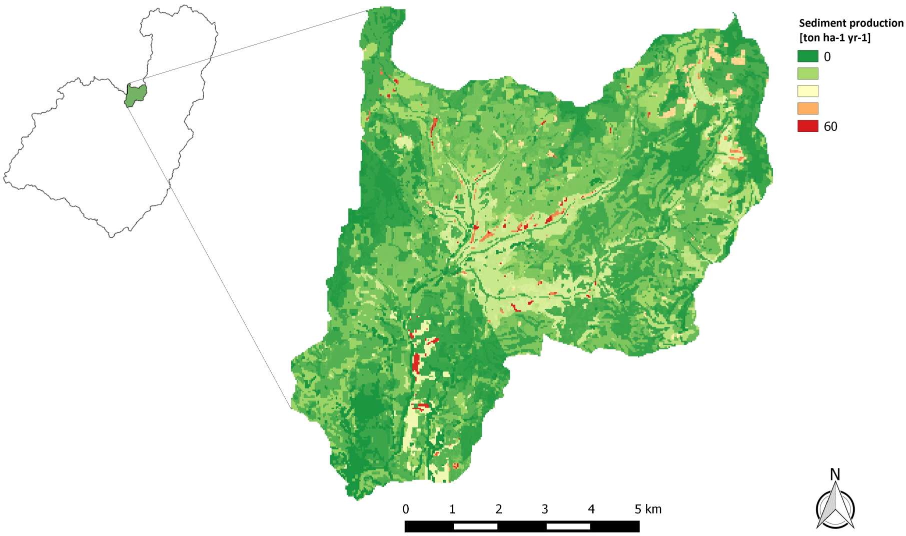

Figure 2.2: Location in Paute and sediment production under initial land use (source: De Keukelaere (2013)) of the Tabacay catchment.

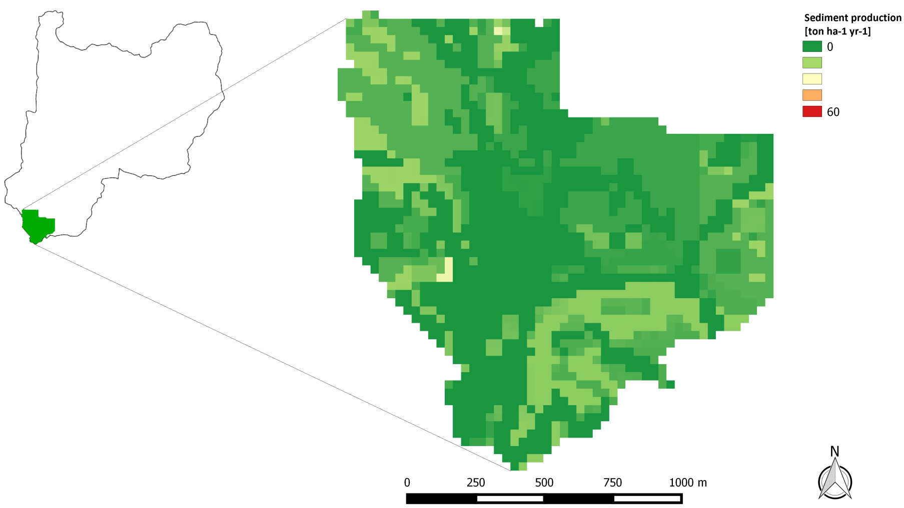

Figure 2.3: Location in the Tabacay catchment and sediment production under initial land use (source: De Keukelaere (2013)) of Tabacay500.

a key issue for sediment flow simulation within CAMF. It refers to the fact that changes in the state of a location can have an impact on the state of neighbouring or even distant locations. In the case of CAMF, spatial interaction refers to the phenomenon that afforestation of a cell leads to changes of its characteristics that in turn affect the amount of sediment flowing from that cell into its downstream cells. Although spatial optimization problems can be formulated as mathematical programming models that can be solved exactly, these models typically require a high amount of computational resources (Fischer and Church (2003); Williams (2002)) while CAMF is meant as a computationally efficient alternative that can be used for optimal site location.

Sediment flow simulation in CAMF is based on an eight-direction (D8) algorithm (O'Callaghan and Mark (1984)) which is an instance of a SFD model based on the assumption that flow follows the steepest-descent route between cells. In other words, flow leaving a cell is assumed to be delivered entirely to the neighbour at lowest altitude.

Input data

The location-specific input data required by CAMF consist of several raster datasets, each of them representing one characteristic of the study area: 1) flow direction (f), 2) initial sediment production (α), 3) initial flow factor (γ), 4) initial retention capacity (σ_1), 5) initial saturation threshold (σ_2), 6) sediment production after afforestation (λ), 7) flow factor after afforestation (ρ), 8) retention capacity after afforestation (σ_3), 9) saturation threshold after afforestation (σ_4), and 10) rivers.

The Flow Direction (f) dataset for CAMF can be seen as a tree that expresses the flow connectivity of cells (Figure 2.4). In such a tree each node corresponds to a raster cell as depicted in Figure 2.4b, with the root node corresponding to the outlet cell and the child-parent relationship representing the sediment flow direction according to the steepest descent pathway. In this representation, child nodes deliver sediment to their parents.

Piecewise linear convex function

The sediment accumulation in each cell corresponds to the total amount of sediment received by that cell from its upslope neighbours (children in the tree representation) plus the amount of sediment produced locally in the cell (α). Considering the non-linear relationship between sediment flow and slope (Postma et al. (2008); Roering et al. (1999)), CAMF makes use of a piecewise linear convex

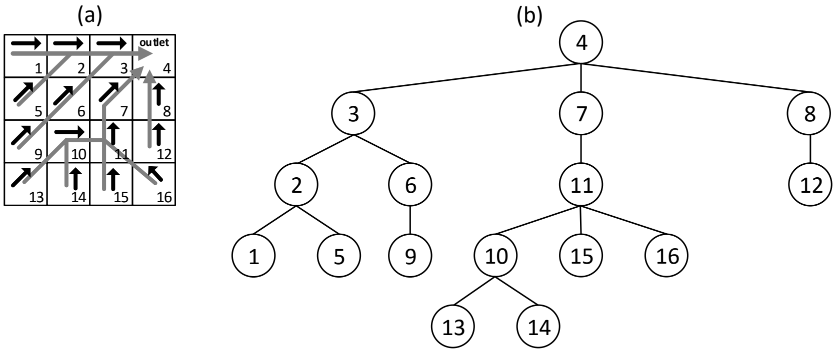

Figure 2.4: (a) Part of a SFD raster dataset (black arrows indicate flow direction, gray arrows denote the tree branches), and (b) its corresponding tree representation (Vanegas et al. (2012)).

function to compute the amount of sediment that leaves a particular cell. As shown by curve a in Figure 2.5, the amount of sediment that leaves a cell (D) is determined by its sediment accumulation and two parameters, namely retention capacity (σ_1) and saturation threshold (σ_2). If the sediment accumulation of the cell is below its retention capacity, no sediment is assumed to leave the cell. If the sediment accumulation of the cell is in the interval between its retention capacity and saturation threshold, a fraction of the sediment accumulation leaves the cell. This fraction is determined by a flow factor parameter (γ). If the sediment accumulation of the cell exceeds the saturation threshold, the exceeding amount of sediment is assumed to be fully delivered to the downslope steepest neighbour cell. The parameters determining the linear piecewise convex function change when a cell is afforested. It is expected that afforestation decreases local sediment production (λ) and flow factor (ρ), whereas it increases the retention capacity (σ_3) and the saturation threshold (σ_4). This is illustrated by curve b in Figure 2.5. Therefore, CAMF requires two sets of input datasets for sediment produced locally, flow factor, retention capacity and saturation threshold: one set representing the catchment under the original land cover (initial situation), and a second set in which every available cell is assumed to be afforested.

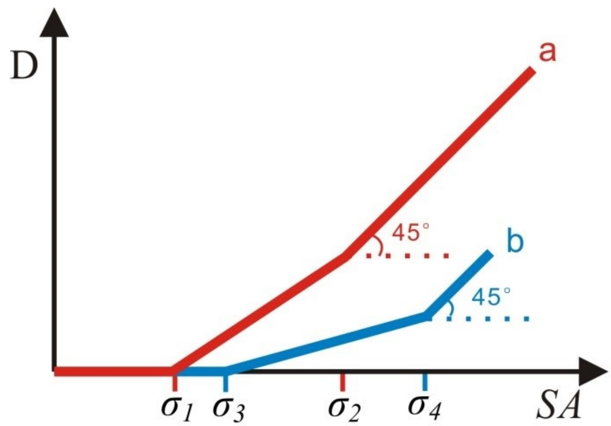

Figure 2.5: Graphs of the piecewise linear convex function used in CAMF for computing the amount of sediment that leaves a cell in case (a) the cell is not covered by forest and (b) the cell is covered by forest. σ_1, σ_2: retention capacity and saturation threshold before afforestation; σ_3, σ_4: retention capacity and saturation threshold after afforestation; D: amount of sediment leaving the cell.

Algorithm

The CAMF algorithm comprises three steps: 1) Computation of sediment accumulation, 2) Cell ranking; and 3) Selection of most suitable cells. Each of these steps is described below.

Step 1: Computation of sediment accumulation The sediment accumulation in leaf nodes in the tree representation (Figure 2.4b) corresponds to the sediment produced locally in each of the corresponding cells, since there is no incoming sediment. To compute the sediment accumulation for all the other cells, they are processed sequentially level by level, from the bottom to the top of the tree representation. The piecewise linear convex function in CAMF is applied to compute the amount of sediment leaving a cell and entering the lowest positioned of its neighbours. The sediment accumulation for each cell amounts to the sediment produced locally plus the incoming sediment. It is the total amount of sediment that is present in a cell at a given point in time. This procedure is continued until the root is reached. The final outcome of this step is the sediment accumulation for every cell in the catchment. Using the sediment accumulation and the piecewise linear convex function of the outlet cell the total annual amount of sediment yield of the catchment (ton yr^{-1}) can be determined.

Step 2: Cell ranking When a single cell is afforested, the change in the sediment produced locally and in the parameter values of its piecewise linear convex function may decrease the amount of sediment delivered to its parent cell. This change in the amount of sediment delivered is propagated from the afforested cell to the outlet. In this way, the decrease in the amount of sediment entering the outlet (sediment yield) is computed. This decrease indicates the sediment reduction potential of the considered cell in case it would be afforested. This process is repeated for every cell in the catchment. Cells are ranked in descending order according to their potential for sediment yield decrease at the outlet.

Step 3: Selection of optimal cells In this step, the cell ranked highest producing the greatest reduction in sediment yield in case it was afforested, is added to the set of selected cells for afforestation (solution set).

By iterating through steps 1 to 3, CAMF adds one or possibly more cell(s) to the solution set until the number of selected cells reaches a user-specified value.

2.2.3 Local CAMF

The variant of CAMF introduced in this section, named local CAMF, aims at avoiding the extra computational cost that considering spatial interaction introduces into the operation of the original algorithm. In local CAMF, the notion of spatial interaction is simply not taken into account and only local information is used to score and rank cells. The decision of neglecting spatial interaction implies that sediment flow simulation is not performed in local CAMF. This made Step 1 of the original CAMF algorithm (Section 2.2.2) no longer necessary. Therefore, local CAMF limits itself to a modified version of Step 2 and to Step 3 of the original algorithm.

In Step 2 of the original CAMF algorithm, a score consisting in the potential for sediment yield reduction of every cell is computed and used as the basis for building a cell ranking. To compute sediment yield reduction, a sediment flow simulation component is integrated in original CAMF. In local CAMF, on the other hand, sediment yield reduction is not considered any more for computing those scores, and, therefore, no sediment flow simulation is performed. Considering the conclusion of Vanegas et al. (2012), who state that, in general, cells with steeper slopes and higher local sediment production are selected during the first iterations of CAMF. Furthermore, taking into account that both slope and local sediment production correspond to information that was readily available for the study regions, those factors were chosen as the basis to compute the required scores that allow to produce a cell ranking in this variant of CAMF. In summary, the score assigned by local CAMF to a cell was computed using Equation 2.1.

$$s_i = w_f f_i + w_e e_i \tag{2.1}$$

where:

- s_i is the score assigned to cell i;
- w_f and w_e are user defined parameters that can take values in the range [0, 1] and indicate the relative importance (weight) assigned to each factor, either slope or sediment production, respectively, with $wf + we = 1$;
- f_i is the normalized (scaled to the range [0, 1]) slope of cell i;
- e_i is the normalized change that would result in local sediment production when cell i was afforested, that is the difference in sediment production between the initial situation and the afforested situation.

Step 3 of the algorithm remains the same, so that the cell or cells at the top of the ranking (maximum value for s_i) are selected. Since both slope and local sediment production values do not change during the execution of this method, local CAMF is not an iterative method and, therefore, all required cells are selected in a single step.

Both the original version of CAMF and local CAMF were implemented in the Java programming language. The layouts of the spatial output were produced using QGIS.

2.2.4 Methodology

Performance measures

The experimental phase consisted in several executions of both original and local CAMF for the three considered geodatabases for solution sizes corresponding to 1, 10, 100 and 1000 cells. Solution size is the term used throughout this chapter to refer to the required number of cells to be selected for afforestation. This parameter is set by the user of CAMF before its execution starts. During each test, several performance factors were recorded, namely:

Sediment yield reduction: Decrease in the sediment yield of a catchment (with respect to the initial situation) when the required number of cells are afforested;

Execution time: CPU time necessary to produce the required output;

Number of iterations: Number of iterations performed by original CAMF to produce the required output;

Spatial coincidence: This is a comparative performance measure applicable only to local CAMF. It uses the output (cells selected for afforestation) produced by original CAMF as a reference. Spatial coincidence was computed as the ratio n_c/n, where n_c is the number of common cells selected by both original and local CAMF, and n is the solution size. Therefore, a spatial coincidence of 1 indicates that both methods selected exactly the same set of cells.

Parameter values

The different input datasets and parameter values used when executing both versions of CAMF are listed in below. The values corresponding to retention capacity and saturation threshold were arbitrarily set in such a way that around half of the available sediment under the original land cover leaves the river catchment in a time unit (year).

- Sediment production (ton ha^{-1} yr^{-1}): available datasets (Section 2.2.1)
- Retention capacity (ton ha^{-1} yr^{-1}):
  - initial (σ_1): Paute: 3, Tabacay: 1.89, Tabacay500: 0.83
  - afforested (σ_3): Paute: 6, Tabacay: 3.78, Tabacay500: 1.67
- Saturation threshold (ton ha^{-1} yr^{-1}):
  - initial (σ_2): Paute: 9, Tabacay: 5.67, Tabacay500: 2.5
  - afforested (σ_4): Paute: 12, Tabacay: 7.56, Tabacay500: 3.33
- Flow factor (-):
  - initial: slope linearly scaled to [0, 1]
  - afforested: initial flow factor divided by 2
- Flow direction (-): computed from DEM, based on D8
- Solution size (number of cells): 1, 10, 100, 1000

2.3 Results and discussion

2.3.1 Original CAMF

The sediment yield reduction values obtained after executing original CAMF are shown in Table 2.1. The sediment yield reduction in case the corresponding number of cells are afforested is shown as an absolute value in the second column and as a percentage with respect to the initial sediment yield in the third column.

It is clear from Table 2.1 that sediment yield reduction values for Tabacay500 and Paute increase almost proportionally with respect to the solution size, which is an indication that at least 100 cells in Tabacay500 and 1000 cells in Paute perform almost equally well when afforested. This is not the case for Tabacay, where such proportionality is evident only when comparing the sediment yield reduction corresponding to solution sizes of 1 and 10. This proportionality is not present when solution sizes of 100 and 1000 cells are considered. This is an indication that in Tabacay the 10 first cells to be selected perform comparatively

Solution size # cells	SYR ton yr^{-1}	% SYR
Tabacay500 (initial SY: 370 ton yr^{-1})		
1	0.498	0.1
10	4.971	1.3
100	46.665	12.6
Tabacay (initial SY: 29075 ton yr^{-1})		
1	3.308	0.01
10	32.171	0.11
100	199.806	0.69
1000	924.975	3.18
Paute (initial SY: 3212203 ton yr^{-1})		
1	14.729	0.0005
10	147.205	0.0046
100	1470.557	0.0458
1000	14675.398	0.4569

Table 2.1: Sediment yield reduction values corresponding to original CAMF. SY = sediment yield; SYR = sediment yield reduction.

well in terms of sediment yield reduction. From the 11th to the 100th and from the 100th to the 1000th selected cells such uniformity in performance is no longer present. It is clear that among those cells there are some for which their performance in terms of sediment yield reduction is considerably lower with respect to cells selected in previous iterations.

Table 2.2 list some performance indicators measured during the tests with original CAMF. The last column in this table shows the average number of cells that are selected at each iteration. The term candidate cells is used to refer to the cells that are available to be selected by CAMF for afforestation, that is, the cells that are under pasture or agriculture, as indicated in Section 2.1.

Except for Tabacay500, execution times seem to increase linearly with respect to solution sizes. This is given by the fact that in almost all cases the number of iterations performed by original CAMF is equal to or slightly smaller than the corresponding solution size. This effect is less noticeable for Tabacay500, for which only very short times are required. In this case, internal details of the implementation and even technical aspects related to the way in which the algorithm is executed by the operating system take a higher relative importance with respect to factors pertaining to the method itself, like simulating sediment

Solution size # cells	CPU time s	# iterations	Cells/iteration
Tabacay500 (total cells 1892, candidate cells 500)			
1	0.015	1	1.00
10	0.046	10	1.00
100	0.109	97	1.03
Tabacay (total cells 68123, candidate cells 26850)			
1	0.234	1	1.00
10	1.872	10	1.00
100	17.799	99	1.01
1000	155.002	927	1.08
Paute (total cells 5616679, candidate cells 1647304)			
1	133.646	1	1.00
10	1311.625	10	1.00
100	11556.164	87	1.15
1000	93484.394	701	1.43

Table 2.2: Performance indicators measured for original CAMF.

flow and building the ranking of cells.

Unexpectedly, in all tests involving solution sizes of 100 and 1000, the number of cells selected per iteration by original CAMF is greater than one. This finding indicates that the probability of more than one cell corresponding to exactly the same maximum sediment yield reduction at a given iteration plays a role in practice. This may be an indication that the function applied to compute the amount of sediment leaving a cell and the way in which sediment flow is simulated play a homogenizing role for the computation of sediment yield reduction values. On the other hand, in all those tests, the average number of cells selected per iteration is still close to one. This characteristic may make CAMF execution times unnecessarily long.

Database size, in terms of number of candidate cells, has a clear impact on execution times. This is explained by the fact that larger database sizes will require more cells to be processed at each iteration. When comparing the execution times obtained for Paute to the corresponding values for Tabacay, a clear proportionality is found. This is not the case when execution times for Tabacay and Tabacay500 are contrasted. This may be the result of (very short) execution times for Tabacay500 being largely influenced by internal, technical aspects of algorithm execution. When considering execution times separately, it can be argued that they start to play a restrictive role for large databases like Paute. Specifically, original CAMF requires more than 3 hours to select 100 cells in Paute, and almost 26 hours for a solution size of 1000 cells. Additionally, it is important to note that the solution sizes tested are rather limited, considering the full size of the database, especially for the case of Paute.

Figure 2.6 shows the output of original CAMF when selecting 1, 10, 100 and 1000 cells in Tabacay.

2.3.2 Local CAMF

The first step conducted when applying local CAMF was to determine sensible values for the relative importance parameters corresponding to slope and sediment (w_f and w_e in Equation 2.1). In this case, a naive trial-and-error approach was used, based on testing different combinations of values for w_f and w_e to score, rank and select cells and assessing the corresponding values of sediment yield reduction produced when the selected cells were afforested. The combinations of parameter values that were tested and their corresponding results in Table 2.3. The values in columns 2 to 6 of Table 2.3 show the ratio between the sediment yield reduction produced by local CAMF when its parameters were set as indicated in the column headers and the sediment yield reduction value produced by original CAMF.

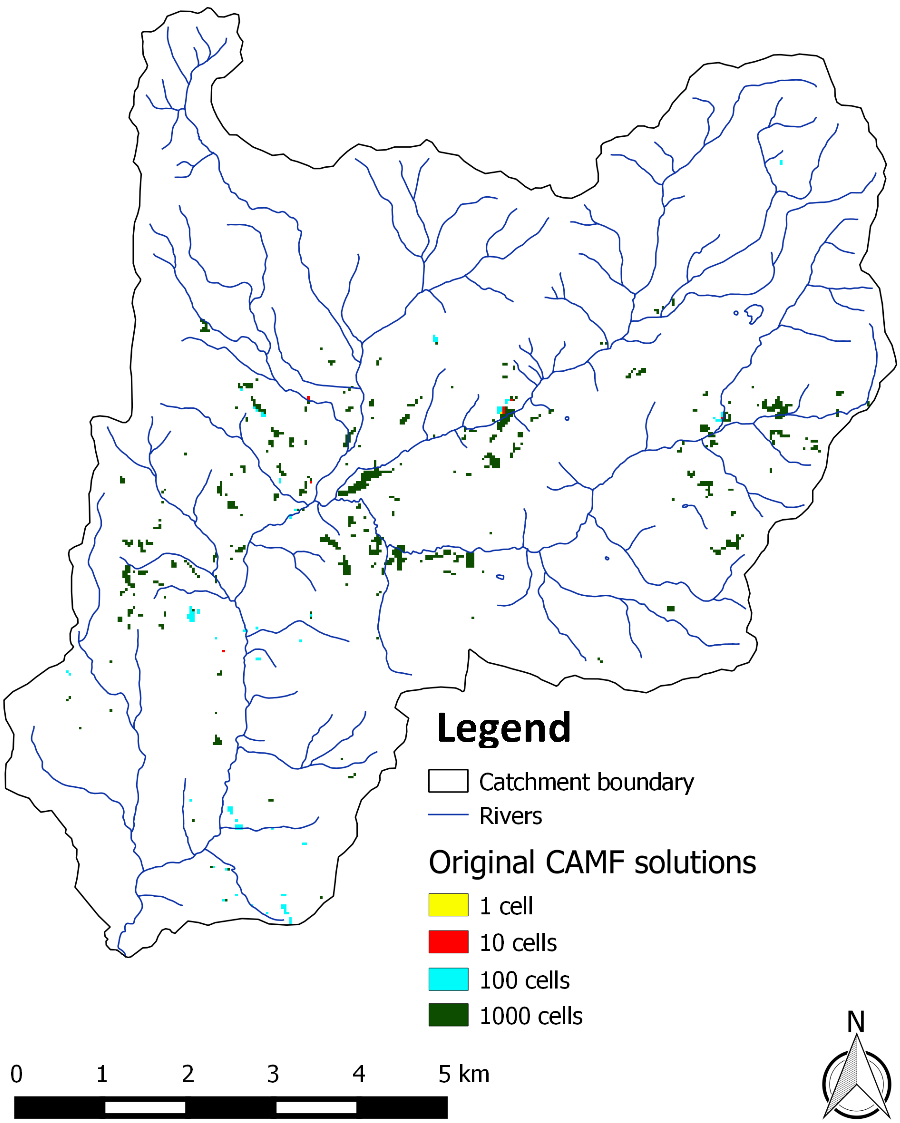

Figure 2.6: Cells selected by original CAMF for a solution sizes of 1, 10, 100 and 1000 cells in Tabacay.

From Table 2.3 we can conclude that setting $w_f = 0.01$ and $w_e = 0.99$ produces the best results from among the tested combinations. This indicates that when assigning a relative importance of 0.99 to local sediment production and 0.01 to slope in Equation 2.1, the results in terms of cells selected for afforestation are virtually the same as for original CAMF. This shows that using almost exclusively local sediment information suffices to obtain almost identical results as original CAMF, and that slope does not play a significant role in this regard. This is indeed opposed to what one would intuitively expect in reality, where cells with steep slopes would be probably preferred to be afforested to reduce sediment transport. Although original CAMF also uses slope information to simulate sediment flow, it seems that given the specific assumptions on which CAMF relies and the characteristics of the input data, the influence of slope on which cells are selected becomes almost negligible, and local sediment production predominates at a large extent in the process of deciding which cells should be afforested.

Considering the results in Table 2.3 w_f and w_e were set to 0.01 and 0.99,

Solution size # cells	$w_f = 0.5$ $w_e = 0.5$	$w_f = 0.75$ $w_e = 0.25$	$w_f = 0.25$ $w_e = 0.75$	$w_f = 0.1$ $w_e = 0.9$	$w_f = 0.01$ $w_e = 0.99$
Tabacay500					
1	0.91	0.90	0.91	0.91	1.00
10	0.95	0.94	0.96	0.98	1.00
100	0.89	0.81	0.99	0.99	0.99
Tabacay					
1	1.00	1.00	1.00	1.00	1.00
10	0.80	0.57	1.00	1.00	1.00
100	0.60	0.42	0.88	0.99	0.99
1000	0.70	0.57	0.88	0.96	0.99
Paute					
1	1.00	1.00	1.00	1.00	1.00
10	1.00	0.99	1.00	1.00	1.00
100	0.99	0.98	1.00	1.00	1.00
1000	0.98	0.93	0.99	1.00	1.00

Table 2.3: Tuning results for the relative importance parameters of local CAMF. Values in columns 2-6 express the ratio between the sediment yield produced by local CAMF (using the parameter values in the column headers) and original CAMF.

respectively, in all tests performed with local CAMF. The resulting sediment yield reduction values are listed in Table 2.4. Column 'SYR fraction' shows the ratio between the absolute sediment yield reduction resulting from local CAMF and original CAMF.

It can be seen from Table 2.4 that local CAMF produces practically the same results as original CAMF. A first interpretation of these results is that spatial interaction does not play a role for the combination of databases and parameter values used during the tests. Considering the values set for the relative importance parameters (w_f and w_e), it can be argued that the local sediment reduction information, that is, the amount in which sediment production would decrease in every cell when afforested, is virtually the only factor that is determining which cells are selected.

Stating that spatial interaction does not play a major role when sediment transport simulation in particular, and off-site criteria in general are involved, may seem counter intuitive. However, this finding can be supported by the fact that relatively limited solution sizes were used during the tests, especially for the cases of Tabacay and Paute. When a limited number of cells are to be

Solution size # cells	SYR ton yr^{-1}	SYR fraction
Tabacay500 (initial SY: 370 ton yr^{-1})		
1	0.498	1.00
10	4.971	1.00
100	46.398	0.99
Tabacay (initial SY: 29075 ton yr^{-1})		
1	3.308	1.00
10	32.171	1.00
100	197.504	0.99
1000	913.868	0.99
Paute (initial SY: 3212203 ton yr^{-1})		
1	14.729	1.00
10	147.205	1.00
100	1470.557	1.00
1000	14675.398	1.00

Table 2.4: Sediment yield reduction values corresponding to local CAMF. SY = sediment yield; SYR = sediment yield reduction.

selected from a large number of candidate cells, it can occur that most selected cells in fact do not interact with each other, which means that they do not share a meaningful segment of their path to the outlet and therefore, changes in the state of one cell do not affect the state of other selected cells. It can be claimed that the river may act as an element that produces interaction among cells, since sediment leaving most cells will eventually reach and be transported by the river to the outlet. However, the river plays the role of a transport channel, that is, it does not really influence the sediment yield attributed to a given cell, or the sediment yield reduction produced when that cell is afforested. This means that all sediment that leaves a cell and reaches the river will be fully transported to the outlet of the catchment, at least for the parameter values used during the tests, especially regarding retention capacity and saturation threshold. It is expected that using larger solution sizes would lead to an increased probability of spatial interaction occurrence among selected cells. In that case, it can be foreseen that local CAMF would produce significantly different results with respect to original CAMF, also this claim is not backed up by the output of local CAMF for Tabacay500.

Table 2.5 lists the performance indicator values measured during the execution of the tests with local CAMF. The column ‘CPU time fraction’ lists the ratio between the execution time of local CAMF with respect to original CAMF.

Regarding execution times of local CAMF, it is clear that they are not influenced by solution size, since once the ranking is built it takes about the same time to select any number of cells from it. On the other hand, execution times are indeed influenced by the database size, since the time spent building the ranking of cells will depend on the number of candidate cells. CPU time fractions show the drastic reduction on execution time that occurs when spatial interaction is left out of consideration and when all cells are selected in a single step, instead of using iterative selection.

2.4 Conclusions

Vanegas et al. (2012) proposed a technique called CAMF with the aim of selecting from a rasterized database representing a river catchment a set of cells to be afforested in order to minimize the sediment yield of the whole catchment. In this chapter an implementation of CAMF was produced and its performance was tested on three databases representing nested river catchments in the southern Andes of Ecuador, with the aim of analyzing the behaviour of CAMF when applied to databases that differ greatly in size. In addition, the influence of the number of cells to be selected on the performance of CAMF was assessed.

In contradiction to what was initially expected, the number of cells selected at each iteration by original CAMF was not exactly 1 in all tests. This indicates that the possibility of two or more cells having exactly the same sediment yield reduction value at a given iteration, although limited, does exist, at least under the implicit assumptions both in the input data and in CAMF itself. However, since the observed deviation from 1 is small or, in other cases, the number of selected cells per iteration is exactly 1, execution times increase almost in direct proportion with respect to solution sizes. Besides solution size, the number of cells comprised in the database has also a clear impact on execution times. This fact allows to conclude that execution time can become a limiting factor for original CAMF, specifically in cases in which it is applied to high resolution databases covering large extents and using large solution sizes. This restriction would be even more apparent in such contexts when several runs of original CAMF are necessary, as it could be the case when performing scenario analysis, or when using original CAMF as a component of an integral model or method that requires to execute it repeatedly in a systematic way.

Solution size # cells	CPU time s	CPU time fraction	Spatial coincidence
Tabacay500 (total cells 1892, candidate cells 500)			
1	0.001	0.000	1.00
10	0.001	0.000	1.00
100	0.015	0.138	0.99
Tabacay (total cells 68123, candidate cells 26850)			
1	0.062	0.265	1.00
10	0.062	0.033	1.00
100	0.046	0.003	0.98
1000	0.093	0.001	0.97
Paute (total cells 5616679, candidate cells 1647304)			
1	3.135	0.023	1.00
10	3.088	0.002	1.00
100	3.634	0.000	0.98
1000	5.834	0.000	0.97

Table 2.5: Performance indicators measured for local CAMF.

A variant of CAMF called local CAMF was also proposed, implemented and tested on the same databases as the original CAMF. This variant uses only local cell information, i.e., sediment reduction and slope, to score and rank cells. Tests using local CAMF produced very similar results with respect to original CAMF outputs, in an almost negligible, constant execution time. One interpretation of this finding may be that for these specific combinations of databases, solution sizes, and parameter values, spatial interaction does not play a role. This observation can be attributed to the fact that solution sizes used in tests are limited when compared to the full database sizes. It is assumed then that, when larger solution sizes are used, the relevance of the spatial interaction role will significantly increase. It is expected that, in such cases, local CAMF would produce different results with respect to original CAMF.

It is clear that the behaviour of CAMF depends heavily on the values set for its parameters. This is especially true for the retention capacities and saturation thresholds for every cell. Values for these and other parameters must be carefully determined, in order for CAMF to reproduce real world phenomena in a valid way. A systematic and scientific sound calibration procedure becomes a requirement in this regard. However, such a procedure involves a more detailed consideration of sediment production and transport, a first approach of which is addressed in the next chapter.

Chapter 3

Determining optimal afforestation patterns with a multiple flow direction model

This chapter is the basis for:
Zhang, K., Estrella, R., Cattrysse, D., Van Orshoven, J. Determining the optimal afforestation pattern in a mountainous catchment with a multiple flow direction heuristic method (in preparation).

Note about the contribution of the author of this dissertation to this article: The author of this dissertation supervised the work that is reported in this article. Frequent discussions about the implementation of CAMF-MFD (the CAMF variant studied in this chapter) and the intermediate outcomes resulting from its application took place between the main author of this article and the author of this dissertation. The author of this dissertation also provided source code corresponding to an implementation of the original version of CAMF, which was used as a basis for implementing the CAMF-MFD algorithm. Finally, the author of this dissertation carried out a thorough review of all the intermediate drafts of the text of this article.

3.1 Introduction

Sediment delivery to rivers reduces channel and reservoir capacity and also leads to water quality problems because of suspended mineral and organic substances (Drzewiecki and Mularz (2008)). These issues are particularly undesirable in regions where the river is used for drinking or irrigation water provision or for electricity production. One measure that has proven to be effective to reduce sediment production and delivery and alleviate the associated water quality problems is afforestation (Costin (1980); Heil et al. (2007); Nearing et al. (2005)). Afforestation is known to decrease runoff and protect the soil surface against the ability of raindrops and runoff to detach and transport sediments (Piégay et al. (2004); Vought et al. (1995)). At the beginning of any afforestation project, one of the main questions to be tackled is to identify the most suitable sites to plant the trees. The discrimination between suitable and unsuitable areas for afforestation typically depends on several criteria adopted by forest planners, which can range from on-site and off-site environmental concerns to maximizing financial profits.

Vanegas et al. (2012) proposed CAMF (Section 2.2.2) as a method to locate sites within a river catchment that should be afforested in order to minimize sediment delivery to the river system. Spatial interaction among cells in the raster datasets representing the catchment is a key issue for sediment flow simulation within CAMF. It refers to the fact that changes in the state of a location can have an impact on the state of neighbouring or even distant locations. In the case of CAMF, spatial interaction refers to the phenomenon that afforestation of a cell leads to changes of its characteristics that in turn affect the amount of sediment flowing from that cell into its downstream cells. Although spatial optimization problems can be formulated as mathematical programming models that can be solved exactly, these models typically require a high amount of computational resources (Fischer and Church (2003); Williams (2002)). CAMF is meant as a less demanding and more scalable alternative to mathematical programming models. CAMF is said to produce similar results as exact methods in general and mathematical programming in particular, while requiring a considerably smaller amount of computational resources, especially execution time (Vanegas et al. (2012)).

Sediment flow simulation in CAMF is based on an eight-direction (D8) algorithm (O'Callaghan and Mark (1984)) which is an instance of a SFD model, which assumes that flow follows the steepest-descent route between cells. In other words, flow leaving a cell is assumed to be delivered entirely to the neighbour at lowest altitude. Despite the advantage of simplicity of SFD models like D8, it has been suggested that they fail to reflect the real nature of surface transport processes (Quinn et al. (1991); Wilson and Gallant (2000)). Quinn et al. (1991)

proposed that Multiple Flow Direction (MFD) models, which are based on the assumption that flow is distributed to one or more of the neighboring downslope cells, can more appropriately represent the sediment flow pathways than SFD. With SFD even a tiny elevation difference between two neighboring cells can have a large effect, since it might determine which cell receives all outgoing flow. Comparatively, small elevation differences have a less important effect in an MFD algorithm, since cells with slight differences in elevation receive about the same amount of flow.

In this chapter we present the reformulation of the original, SFD version of CAMF (CAMF-SFD) into its MFD variant CAMF-MFD, considering the transition from Single to Multiple Flow Direction as a potential improvement for this method. A model inversion calibration was performed to obtain realistic values for the location specific parameters for CAMF-MFD when applied to the medium size Tabacay river catchment (Section 2.2.1) represented by a raster of more than 165000 cells. Vanegas et al. (2012) do not report on any calibration of the CAMF-SFD parameters, since the aim of that work was testing and applying CAMF-SFD to sample data. Finally, the results of applying the calibrated CAMF-MFD method to the Tabacay river catchment (Figure 2.2) for various afforestation scenarios are presented and discussed.

3.2 CAMF-MFD

Several MFD models were explored as potential replacements of the SFD model in the sediment flow simulation component of CAMF-SFD (Section 2.2.2), e.g. the D∞ algorithm proposed by Tarboton (1997), the Digital Elevation MOdel Networks (DEMON) proposed by Costa-Cabral and Burges (1994) and the Fractional Deterministic Eight-Neighbour (FD8) model proposed by Quinn et al. (1991). FD8 was chosen since it is based on the more intuitive assumption that the flow leaving a particular cell can be delivered to all the neighbours located at lower altitudes Quinn et al. (1991). The fraction of sediment delivered to each downslope neighboring cell computed with FD8 is given by Equation 3.1.

$$f_{ij} = \frac{\tan g_{ij} l_{ij}}{\sum_{k=1}^{n} \tan g_{ik} l_{ik}} \tag{3.1}$$

where:

- f_{ij} is the fraction of sediment delivered from central cell i to its neighbour cell j;

- g_{ij} is the gradient between central cell i and neighbour cell j;
- l_{ij} is the contour length, which is equal to 0.5 times the cell size for cardinal neighbours and 0.354 times the pixel size for diagonal neighbours. The resolution of the datasets used in this chapter is 20x20 m, therefore l_{ij} takes a value of 10 for cardinal neighbours and a value of 7.08 for diagonal neighbours. Contour lengths are used in this equation as weighting parameters, such that cardinal neighbours receive a higher fraction of the outgoing sediment than diagonal neighbours;
- n is the number of downslope neighbour cells.

Figure 3.1a shows a schematic representation of an MFD graph, in which the directed edges represent sediment delivery. The corresponding adjacency list structure is shown in Figure 3.1b. Each row in the adjacency list corresponds to a cell. For each cell in the list, two related lists are necessary. The one to the left contains the cells from which it receives sediment and the one to the right contains cells that receive sediment from it. For example, cell 1 receives sediment from cells 5 and 6, which are regarded as its 'ancestors'. Sediment leaving cell 1 is delivered to cell 2, which is its 'successor'.

The only difference between the algorithms of CAMF-SFD and CAMF-MFD is how sediment flow over the catchment is simulated. In particular the way in which the sediment accumulation and the potential sediment delivery reduction of each cell are computed are different in CAMF-MFD with respect to CAMF-SFD.

To compute the sediment accumulation in CAMF-MFD, all cells in the adjacency list (Figure 3.1b) are sorted using the topological sorting algorithm proposed by Kahn (1962). For each pair of cells ('ancestor', 'successor') in the adjacency list, the cell 'ancestor' will always appear before its 'successor' in the resulting sorted list (Figure 3.1c). The computation of sediment accumulation is then performed sequentially cell by cell following the order given by the sorted list, from left to right.

Equation 3.1 is used to compute the fraction (f_{ij}) of accumulated sediment that flows from cell i to each of its downslope neighbour cells j. This value is then used in Equation 3.2 to compute the amount of sediment transported from cell i to j.

$$d_{ij} = d_i f_{ij} \tag{3.2}$$

where:

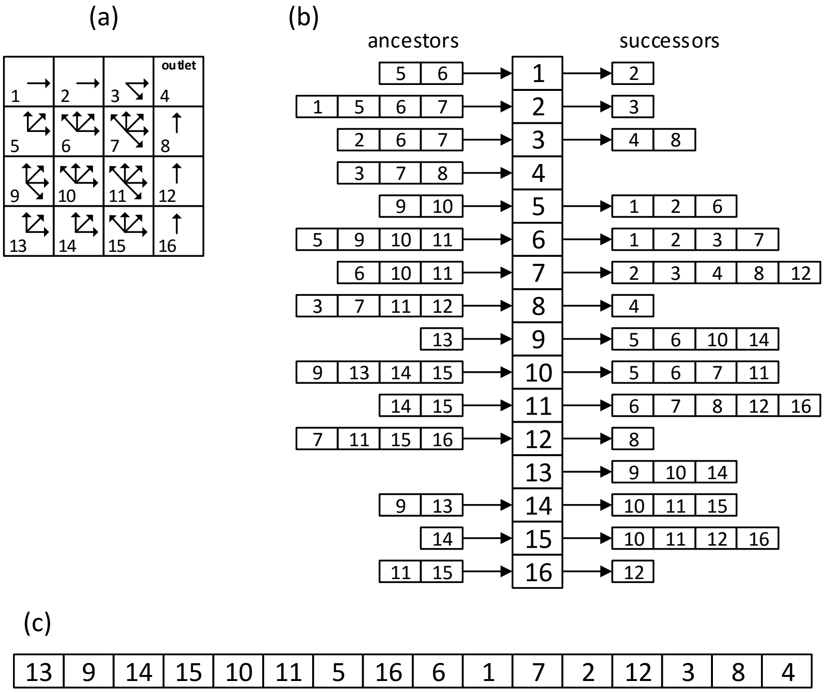

Figure 3.1: Representation of flow direction in CAMF-MFD: (a) part of a MFD raster dataset; (b) adjacent list representation; (c) a topologically sorted list of cells.

- d_{ij} is the amount of sediment transported from cell i to j;
- d_j is the total amount of sediment that leaves cell i. This value is computed using the same piecewise linear convex function as for CAMF-SFD;
- f_{ij} is the fraction of sediment transported from cell i to j.

3.2.1 Input datasets and parameter calibration

Since CAMF-MFD is based on the assumption that every cell can deliver its sediment to multiple neighbours, its flow direction dataset (f) is represented by a multiple direction table. Besides an identifier, this table stores eight values for each cell: one value for each possible flow direction. A value of 1 indicates that there exists flow in the corresponding direction, and 0 indicates otherwise.

The initial sediment production (α) dataset corresponds to the amount of sediment produced locally in each cell under the original land cover. For this study, this dataset was computed by applying the Revised Universal Soil Loss Equation (RUSLE, Equation 3.3) proposed by Renard et al. (1991). The values for the parameters of RUSLE corresponding to the support practice factor (P) and rainfall erosivity (R) were obtained from Cisneros (1999), who reports a value of 1 for P and a value of 1599 MJ mm ha^{-1} h^{-1} yr^{-1} for R in the case of Tabacay. The L*S map was generated using WaTEM/SEDEM. WaTEM/SEDEM is a spatially distributed soil erosion and sediment delivery model that combines an updated version of the RUSLE with a MFD approach to calculate mean annual soil loss values, considering the mean annual transport capacity T_C (the maximum amount of soil that can leave a cell) to be directly proportional to the potential rill erosion (Van Oost et al. (2000); Van Rompaey et al. (2001); Verstraeten et al. (2002)). The equation introduced by Wischmeier and Smith (1978) was used to calculate the K-factor from soil granulometric fractions. It produced the results listed in Table 3.1. The C-factor was determined by assigning the C-values presented in Table 3.2 to the land cover classes shown in Figure 3.2. The values in Table 3.2 were retrieved from De Keukelaere (2013).

$$E = RKLSCP \tag{3.3}$$

where:

- E is the annual soil loss (ton ha^{-1} yr^{-1});
- R is the rainfall erosivity factor (MJ mm ha^{-1} h^{-1} yr^{-1});
- K is the soil erodibility factor (ton h MJ^{-1} mm^{-1});
- L is the slope length factor (-);
- S is the slope steepness factor (-);
- C is the cover factor (-);
- P is the support practice factor (-).

The LUT páramo refers to neotropical andean wetlands normally covered by grassland and located between 3500 and 5000 m asl in the northern Andes (Buytaert et al. (2006)).

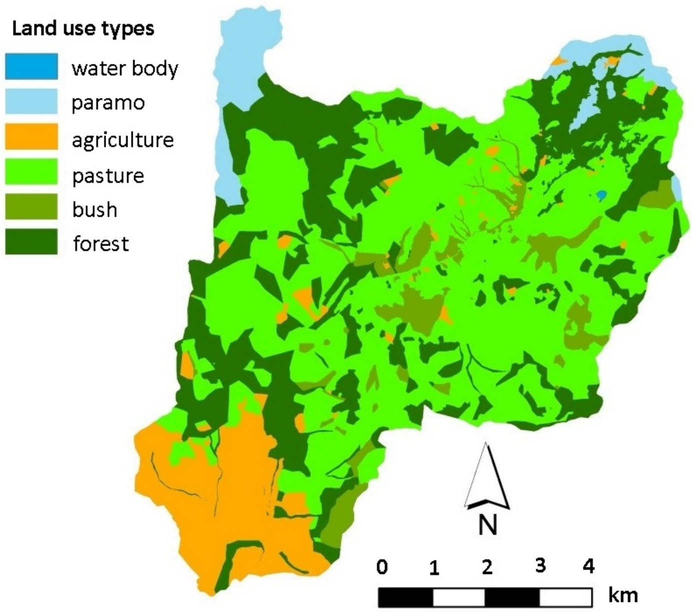

Figure 3.2: Land cover map of the Tabacay catchment (source: PROMAS (2005)).

A normalized slope dataset derived from the DEM was used as the initial flow factor (γ). To compute the other datasets, namely initial retention

Soil type	Soil texture	K factor ton h MJ^{-1} mm^{-1}
Umbric Leptosol	Sandy clay loam	0.0397
Umbric Andosol	Loam	0.0373
Dystric Cambisol	Loam	0.0376
Ferralic Cambisol	Clay loam	0.0253
Calcaric Regosol	Sandy loam	0.0452
Calcaric Cambisol	Sandy clay loam	0.0290
Eutric Cambisol	Silt loam	0.0518
Eutric Regosol	Sandy loam	0.0521

Table 3.1: Soil erodibility factor values for each soil type in Tabacay (computed according to Wischmeier and Smith (1978)).

capacity (σ_1), initial saturation threshold (σ_2), sediment production after afforestation (λ), retention capacity after afforestation (σ_3), saturation threshold after afforestation (σ_4) and flow factor after afforestation (ρ), a systematic calibration procedure was applied. Considering that the study region was represented by raster datasets containing a relatively large number of cells, and that for each cell several parameter values must be defined, two assumptions were made in order to avoid excessive complexity:

Assumption 1: The initial retention capacity (σ_1), initial saturation threshold (σ_2), the retention capacity after afforestation (σ_3) and the saturation threshold after afforestation (σ_4) for a given cell were assumed to be fractions (p_1, p_2, p_3 and p_4, respectively) of its initial sediment production (α), which was obtained by applying the RUSLE. Since the retention capacity of a given cell should be smaller than its saturation threshold, the range of p_1 is $[0, p_2)$ and the range of p_2 is $(p_1, 1]$. Considering that the retention capacity and saturation threshold of a cell are assumed to increase after afforestation, the range of p_3 is (p_1, p_4) and the range of p_4 is $(p_2, 1]$.

Assumption 2: The sediment production after afforestation (λ) is a fraction (p_5) of the initial sediment production (α). The flow factor after afforestation (ρ) is a fraction (p_6) of the initial flow factor (γ). Both p_5 and p_6 are in the range $[0, 1]$, since the sediment production and the flow factor value of an afforested cell are expected not to exceed their corresponding initial values.

With these assumptions in place, the calibration of σ_1, σ_2 (initial situation), σ_3, σ_4, λ and ρ (afforested situation) for each cell is reduced to the computation of two parameters (p_1 and p_2) for the initial situation and four parameters (p_3, p_4, p_5 and p_6) for the afforested situation.

LUT	C factor [-]
Water body	0.000
Páramo	0.003
Agriculture	0.200
Pasture	0.003
Bush	0.003
Forest	0.001

Table 3.2: Cover factor values assigned to each land cover type in Tabacay (source: De Keukelaere (2013)).

To tune the six parameters described above, the model inversion method was applied given its simplicity and adaptability to nearly all model types as suggested by Gao and Lesht (1997); Privette et al. (1997); Weiss et al. (2000); Zarco-Tejada et al. (2001). The procedure consisted in iteratively running CAMF-MFD using different sets of parameter values until a parameter set is found, such that it leads to the smallest difference between the reference results and those produced by CAMF-MFD. Specifically, the following steps were performed:

Step 1: Compute the reference sediment yield value;

Step 2: Initially, set each undefined parameter in CAMF-MFD is to its minimum value. Call this parameter setting S_{opt}; this value will be successively increased by a small step (0.01) in subsequent iterations until its maximum value is reached;

Step 3: Calculate the sediment yield value using CAMF-MFD and S_{opt}; compute the difference between this value and the reference sediment yield;

Step 4: Set the CAMF-MFD parameters to a new set of values (S_1) by picking different values from parameters' possible value set generated in Step 2. If the difference in sediment yield between the reference and the value corresponding to S_1 is less than the difference corresponding to S_{opt}, set S_{opt} to S_1;

Step 5: Iterate Step 4 until all possible combinations of the parameters' values are considered.

The sediment yield value expresses the amount of sediment that leaves a river catchment through its outlet. Sediment yield corresponds to the off-site criterion that is minimized in CAMF and any of its variants. To compute the sediment yield, sediment flow over the catchment is simulated using the initial sediment production dataset as a basis. The reference sediment yield value was computed in this case using the WaTEM/SEDEM model (version 2004). WaTEM/SEDEM was selected as the reference model since it has been validated and widely applied to catchments under various environmental conditions (Van Rompaey et al. (2005, 2003, 2001)). The input data required for running WaTEM/SEDEM in Tabacay are detailed below:

- Digital Elevation Model (DEM);
- Land cover dataset, derived by classifying the land cover map according to the land use identifiers required by WaTEM/SEDEM;

- C-factor dataset, generated using the land cover map and the C-values in Table 3.2;
- K-factor dataset, produced using the soil type map and the K-values in Table 3.1;
- R-factor = 1599 MJ mm ha^{-1} h^{-1} yr^{-1} (Cisneros (1999));

All the other input parameters, including the transport capacity coefficient, were set to the default values.

WaTEM/SEDEM was executed for two scenarios: the initial situation (pre-existing forest covers about 27% of the whole catchment); and the fully afforested situation (all cells, except rivers, are covered by forest). This procedure resulted in two reference sediment yield values. The optimal values for p_1 and p_2 were tuned using the sediment yield value corresponding to the initial situation, while the values for p_3, p_4, p_5 and p_6 were calibrated with respect to the sediment yield computed for the fully afforested scenario.

The procedure followed to determine the values for both the computed parameters (initial sediment production, initial flow factor and flow direction) and the calibrated parameters (thresholds, sediment production after afforestation, flow factor after afforestation) is summarized in the flowchart displayed in Figure 3.3.

3.2.2 Accuracy assessment

To evaluate the accuracy of CAMF-MFD for sediment transport simulation, nine afforestation scenarios were considered. The ith scenario was defined as the afforestation of i*10% of all available cells. A new land cover map corresponding to each scenario was generated considering the cells selected for afforestation by CAMF-MFD. This new land cover map and its corresponding new cover factor map were then used as input for WaTEM/SEDEM to compute the reference sediment yield value pertaining to each scenario. The accuracy of CAMF-MFD was estimated using the Relative Root Mean Square Deviation (RRMSD, Equation 3.4).

$$RRMSD = \frac{\sqrt{\frac{1}{n}\sum_i^n (W_i - N_i)^2}}{\frac{1}{n}\sum_i^n W_i} \tag{3.4}$$

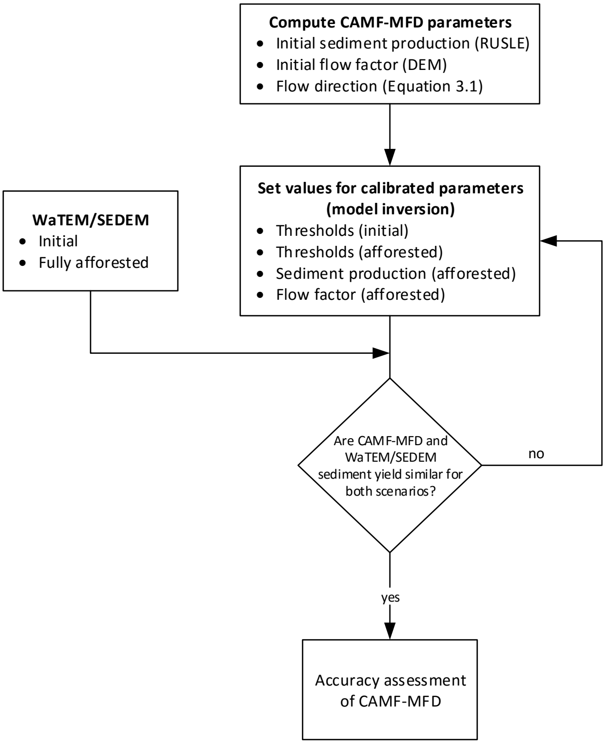

Figure 3.3: Flowchart of the procedure to compute and calibrate the parameters of CAMF-MFD.

where W_i and N_i are the sediment yield values resulting from WaTEM/SEDEM and CAMF-MFD, respectively, and n is the total number of afforestation scenarios that were considered.

The accuracy assessment described above is illustrated as a flowchart in Figure 3.4.

3.2.3 Software tools

CAMF-MFD was implemented in the Java programming language. The MFD table was stored using the PostgreSQL free database management software.

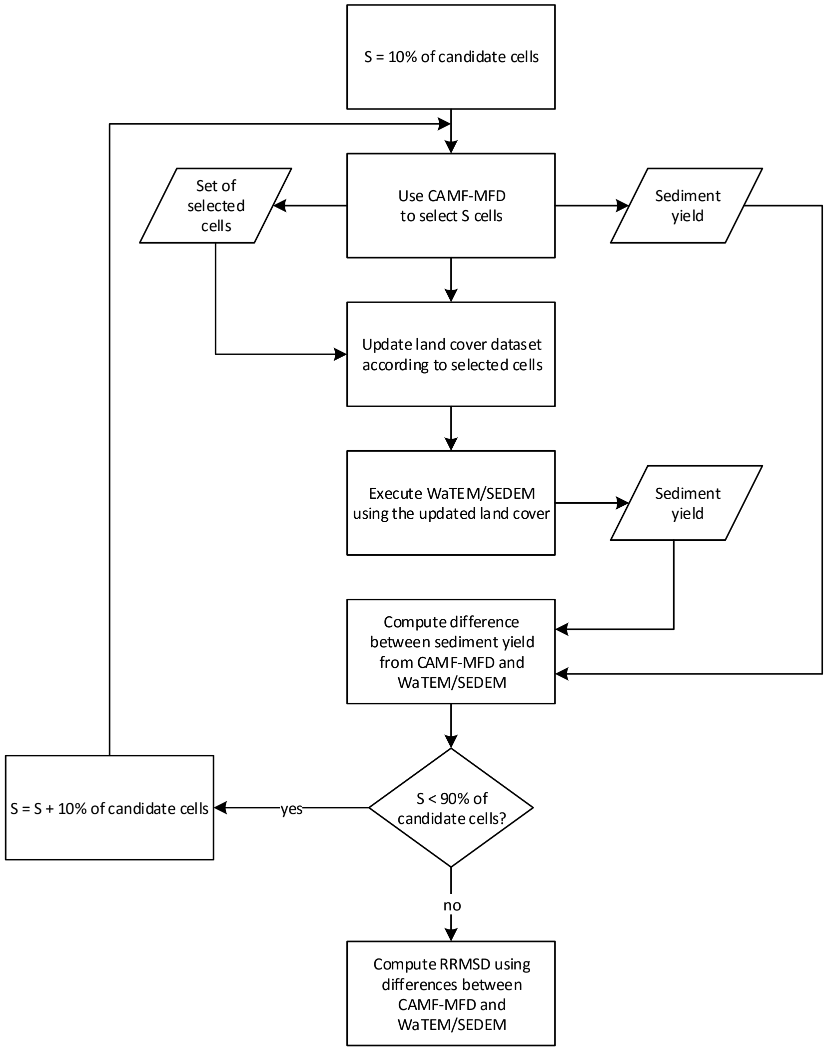

Figure 3.4: Flowchart of the procedure to assess the accuracy of CAMF-MFD.

The pre-processing of the input data for WaTEM/SEDEM was performed using the IDRISI geographic information software.

3.3 Results

3.3.1 Calibration of parameters

Table 3.3 shows the set of parameter values that were obtained from the application of the model inversion method for the input datasets of CAMF-MFD. The corresponding sediment yield values produced by CAMF-MFD and WaTEM/SEDEM are also reported. The datasets retention capacity and saturation threshold for the initial situation, and sediment production, retention capacity, saturation threshold and flow factor for the afforested situation were derived applying these values and Assumptions 1 and 2.

3.3.2 Accuracy assessment

The sediment yield values produced by running CAMF-MFD in nine afforestation scenarios, as well as the corresponding reference values are shown in Table 3.4. Since the total number of cells that can be afforested within the Tabacay catchment is 117291, the number of cells afforested in the ith scenario equals to i * 11729. An RRMSD of 0.04 was obtained for the tests corresponding to all these scenarios.

Scenario	Parameter values	Sediment yield [ton yr^{-1}] CAMF-MFD	WaTEM/SEDEM
Initial situation	$p_1 = 0.37$, $p_2 = 0.96$	996.54	996.43
Fully afforested	$p_3 = 0.37$, $p_4 = 0.96$ $p_5 = 0.37$, $p_6 = 0.96$	46.75	46.72

Table 3.3: Values for the parameters of CAMF-MFD obtained from the calibration procedure and corresponding sediment yield values computed by CAMF-MFD and WaTEM/SEDEM (reference).

3.4 Discussion

As shown in Table 3.4, the sediment yield computed by both CAMF-MFD and WaTEM/SEDEM decreases for every additional 10% of afforested cells. This result is in agreement with previous works that have proposed afforestation as an effective means to reduce sediment production and delivery. It is also apparent in Table 3.4 that CAMF-MFD consistently estimated slightly higher sediment yield values than the reference model in all the afforestation scenarios. This overestimation can be attributed to the field boundary effect incorporated in the WaTEM/SEDEM, which assumes that a fraction of sediment is trapped at a field boundary when the LUT at both sides of the boundary differ (Van Oost et al. (2000)). When a cell is selected to be afforested, it is likely that its LUT becomes different from those of its neighbours. Hence, new field boundaries with different LUTs at both sides are created, which results in a higher level of sediment retained in these boundaries. CAMF-MFD, on the other hand, does not take the additional sediment retention due to the field boundary effect into account.

Figure 3.5 shows the absolute amount of sediment reduction corresponding to the number of cells assumed to be afforested in each scenario. The sediment reductions of CAMF-MFD and WaTEM/SEDEM were calculated by subtracting the sediment yield in the particular scenario from the value of the initial situation computed by each of these models. Although the amount of sediment yield reduction increases, the reduction rate decreases with the number of afforested cells in both models. This is particularly evident from scenario 3 onwards. The difference among the sediment yield reduction values corresponding to the

Scenario	Cells to be afforested	Sediment yield [ton yr^{-1}]	
		CAMF-MFD	WaTEM/SEDEM
1	11729	171.23	164.87
2	23458	112.50	108.81
3	35187	89.83	87.11
4	46916	74.62	72.69
5	58645	63.56	62.42
6	70374	55.43	54.14
7	82103	49.99	48.91
8	93832	47.34	46.71
9	105561	46.76	46.48

Table 3.4: Results produced by CAMF-MFD and reference values computed by WaTEM/SEDEM for each afforestation scenario in the Tabacay catchment.

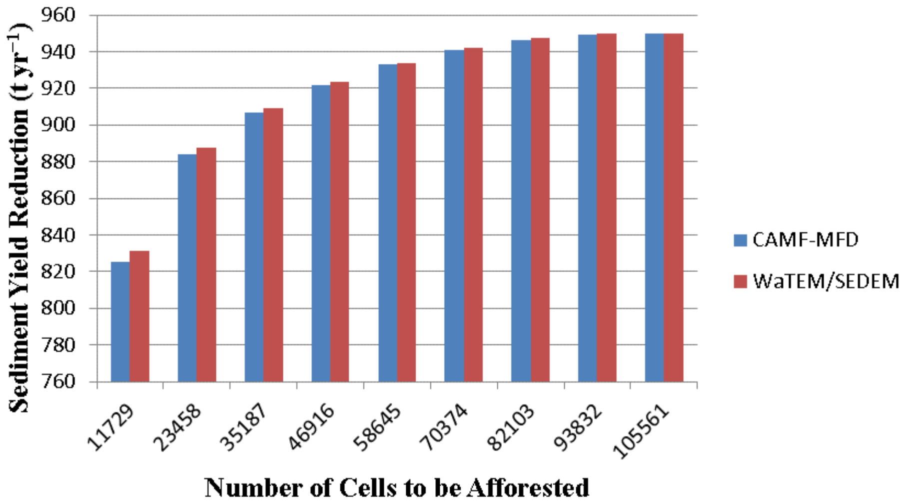

Figure 3.5: Evolution of the sediment yield reduction values with the number of cells to be afforested corresponding to CAMF-MFD and WaTEM/SEDEM (reference).

scenarios requiring the afforestation of relatively large areas (e.g., Scenario 6, 7, 8, 9) is almost negligible. This result confirms that CAMF-MFD selects first the cells leading to the highest sediment yield reduction.

When Figures 3.6 and 3.7 are contrasted, it is clear that most cells selected to be afforested in Scenario 1 and Scenario 2 are concentrated in areas with high sediment production. As a matter of fact, the percentage of cells with sediment production higher than 200 ton ha^{-1} yr^{-1} selected as part of the optimal solution is 99.5% (11672 out of 11729) for Scenario 1. This value is relatively high, especially considering that the percentage of such cells among the whole catchment is only 8% (13349 out of 165983). For Scenario 2 the solution includes all such cells. This observation suggests that the most important factor for a cell to be included in the solution is its initial sediment production, namely, its erosion level computed with the RUSLE. On the other hand, only 12% (1415 out of 11729) of optimal cells in Scenario 1 corresponds to areas with a slope (initial flow factor = normalized slope) higher than 26 degrees, while the percentage of such areas in catchment is 28% (45974 out of 165983). This result indicates that slope is less decisive for cell selection in CAMF-MFD than initial sediment production. Cells with high slope start to appear in the solution from Scenario 2 onwards. This behaviour was also reported for CAMF-SFD (Vanegas et al. (2012)).

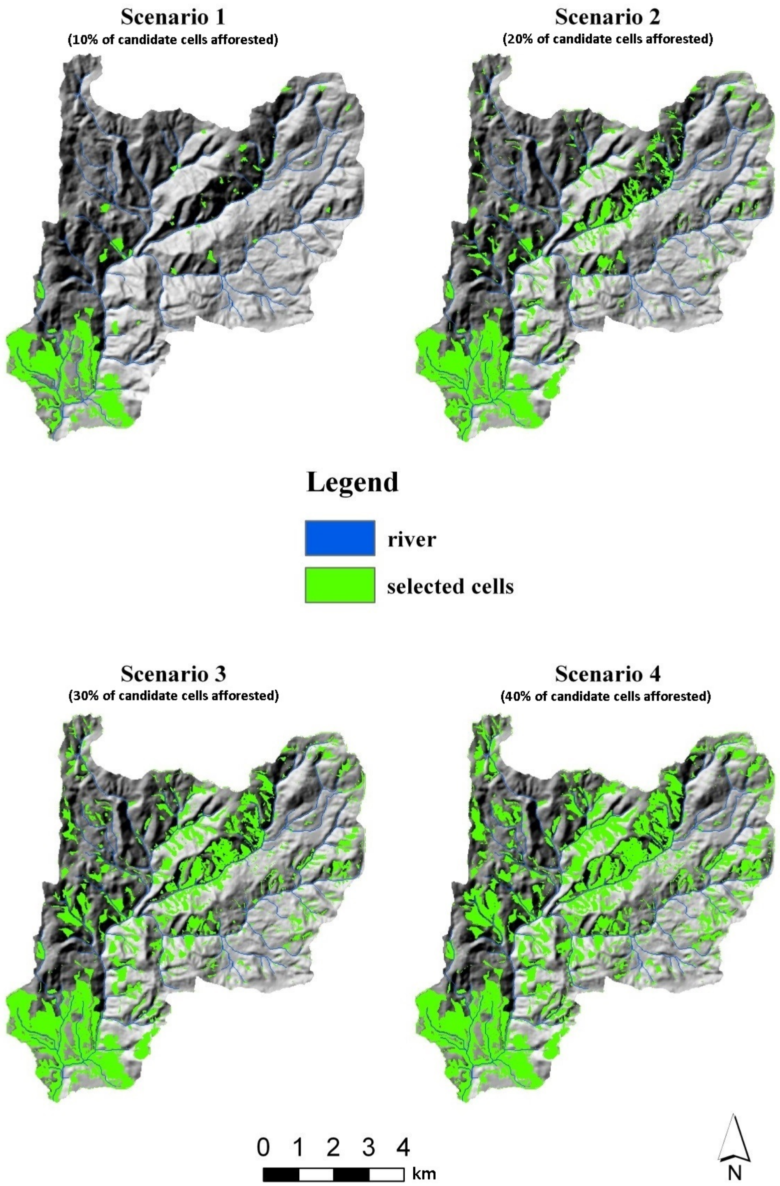

Figure 3.6: Areas selected for afforestation by CAMF-MFD for Scenarios 1-4.

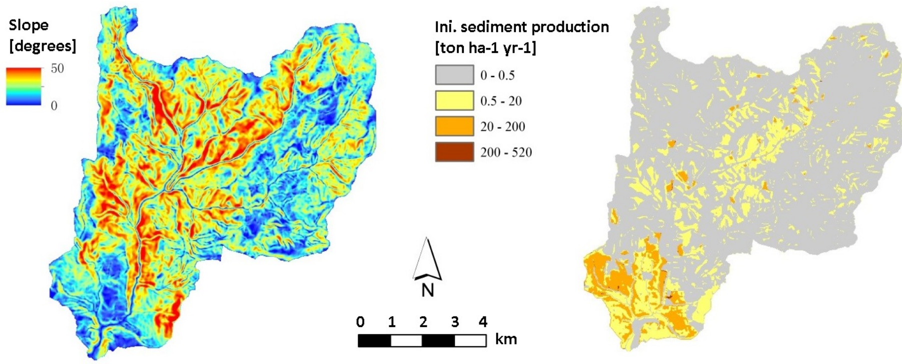

Figure 3.7: Slope and initial sediment production of the Tabacay catchment.

3.5 Conclusions

A MFD variant of CAMF (Vanegas et al. (2012)), called CAMF-MFD, is introduced. It is meant to support the selection of optimal sites for afforestation where minimizing sediment delivery to the river system is the goal. CAMF-MFD encompasses the MFD FD8 algorithm in its sediment flow simulation component. All location specific input data for CAMF-MFD were generated from a DEM and from the output of a spatially distributed RUSLE-based erosion model by means of the inverse calibration of six parameters.

The results of applying CAMF-MFD in a set of afforestation scenarios for the Tabacay catchment (southern Andes of Ecuador) demonstrated that CAMF-MFD is capable of iteratively selecting those cells for which the marginal contribution of afforestation to the sediment yield reduction is highest. The major characteristic of the selected cells is that they are highly prone to soil loss according to the RUSLE. Being located on a steep slope is ranked second. The absolute sediment yield computed by CAMF-MFD is comparable to the one computed by a reference model (WaTEM/SEDEM) for the same afforestation pattern.

One of the limitations of CAMF-MFD is that it does not consider temporal aspects. For instance, the impact of tree growth on the sediment processes occurring in a river catchment is not taken into account in the formulation of this method. This impact could be reflected on, for example, the fact that the sediment retention capacity of areas within the catchment increase gradually during the lifespan of trees in a forest. In addition to temporal aspects, the consideration of budgetary or other resource constraints in afforestation planning are topics that need further attention when techniques are developed to search for locations with a view to maximize off-site benefits.

Chapter 4

Locating afforestation sites to optimize multiple on-site criteria

This chapter is based on:
Estrella, R., Delabastita, W., Wijffels, A., Cattrysse, D., Van Orshoven, J. (2014). Comparison of multicriteria decision making methods for selection of afforestation sites. *Revue Internationale de Géomatique*, 24 (2), 143-157.

4.1 Introduction

The applicability of MCDM methods to natural resources management have been widely demonstrated (e.g., Mendoza and Martins (2006); Ananda and Herath (2009); Ramanathan (2001)). The question of selecting the most appropriate method given an instance of a decision problem is not always trivial. In this chapter an exploration of the literature on MCDM methods is performed, with the aim of determining candidate methods that can be useful to solve the problems addressed in subsequent chapters of this dissertation. The applicability and usefulness of the studied methods is tested for solving an instance of a site location for afforestation problem. Furthermore, the consistency in the results produced by those methods is analyzed and discussed.

This chapter starts with a brief conceptual introduction on multi-criteria decision analysis in Section 4.2. Next the application of six different MCDM methods is demonstrated. The results of all these MCDM methods are compared and discussed. To select the MCDM methods studied in this chapter, the most frequently referred methods were chosen from a literature exploration. The selected MCDM methods are ELECTRE III (Roy (1991)), PROMETHEE II (Brans and Mareschal (2005)), the Analytic Hierarchy Process (AHP, Saaty (1977)), Compromise Programming (Zeleny (1973); Yu (1973)), Stochastic Multi-criteria Acceptability Analysis 2 (SMAA-2, Lahdelma and Salminen (2001)) and the Iterative Ideal Point Thresholding method (IIPT, Gilliams et al. (2005b); De Meyer et al. (2013)). These MCDM methods were applied to the database covering the Tabacay river catchment described in Section 4.3.1. The problem solved is expressed in the question "Where (within the Tabacay catchment) should a pine forest be planted in order to obtain an optimal 'per land unit' land performance 30 years later?"

4.2 Fundamentals of multi-criteria decision analysis

Romero and Rehman (2003) defines the term decision making as the selection made by one or more individuals of the best alternative(s) among the feasible solutions. The individual or group of individuals who are responsible of making the decision is referred to as the decision maker, while the feasible solutions are the decision alternatives that satisfy the constraints of the decision problem. These constraints are dictated by the limited availability of resources (Winston and Goldberg (2004)).

From a high level perspective, a decision making process involves two phases: a) determination of the set of feasible solutions, which can be finite or infinite, and b) selection of the best alternative among the feasible solutions (Malczewski (1999)).

The definition of the best alternative is given by a criterion function, which is a function that associates a real number, or score, to every feasible solution (Opricovic and Tzeng (2004)). Other terms used to refer to this function are utility function, value function or, in the context of mathematical programming, objective function. This function is used to determine the best solution from the feasible set while appropriately reflecting the preferences of the decision maker (Keeney and Raiffa (1993)).

In realistic scenarios the selection of the best solution is hardly based on a single criterion. A typical decision problem requires the simultaneous consideration of several conflicting criteria that requires to find a trade-off solution. This process is known as multi-criteria decision making (Kiker et al. (2005)).

The attributes in a multi-criteria decision problem are the factors that are considered for assessing the set of feasible solutions when making a decision. It must be possible to measure attribute values objectively, that is, there cannot exist any influence coming from the decision maker's preferences on the attribute values (Lahdelma et al. (2000); Zanakis et al. (1998)). Examples of attributes in the context of environmental performance of land can be sediment production or biomass carbon sequestration per land unit under a given LUT.

An objective expresses the improvement direction of an attribute. This improvement direction typically is one of two alternatives: "the more the better" or "the less the better" (Doumpos and Zopounidis (2002)). In the former case the decision maker wants to maximize the attribute, while in the latter the attribute is to be minimized. Examples of objectives are "minimize the amount of sediment produced" and "maximize the amount of carbon stored in the soil".

4.3 Materials and methods

4.3.1 Study region and dataset

The study region in this chapter is the catchment of the river Tabacay (Section 2.2.1). To compute the dataset used as input for the methods discussed in this chapter, the Tabacay catchment was stratified into 20 land units. Each of these land units was defined as an area with homogeneous land performance according to the attributes listed in Table 4.1.

The values of these performance attributes, which were retrieved from Wijffels and Van Orshoven (2009), represent the expected performance of each land

Attribute	Short form	Unit	Optimization
Runoff production	Runoff	mm ha^{-1} yr^{-1}	Minimization
Sediment production	Sediment	ton ha^{-1} yr^{-1}	Minimization
Soil Organic Carbon	SOC	ton ha^{-1} 30cm^{-1}	Maximization
Biomass Organic Carbon	BOC	ton ha^{-1}	Maximization
Monetary income	Income	USD ha^{-1}	Maximization

Table 4.1: Land performance attributes considered as decision criteria.

unit after 30 years of being covered by pine forest. Wijffels and Van Orshoven (2009) computed runoff production values using the rational formula (Rossmiller (1982)) and sediment production values with the formula introduced in (Molina et al. (2008)). To compute the carbon stock in soil, Wijffels and Van Orshoven (2009) report the use of previously available data complemented with field measurements.

The computation of the amount of carbon stored in biomass was based on the assumption that 50% of the biomass corresponds to organic carbon (FAO (1993)). Wijffels and Van Orshoven (2009) determined tree growth curves and used them as a basis to define four site quality classes. Mean values for the amount of biomass were derived from dendrometric parameters like diameter at breast height and total tree height for each site quality class. The values for these parameters were either measured on the field or retrieved from available databases in Ecuadorian governmental institutions of the forestry sector.

The values for monetary income were computed applying Equation 4.1 to each land unit.

$$m = m_p - m_i \tag{4.1}$$

where m is the monetary income value, m_p is the amount of money that the owner of the land would obtain from a pine forest after 30 years, and m_i is the amount of money that would be earned in case the initial LUT (iLUT, e.g., agriculture) is kept in place for the same period of 30 years. For pine forest only the income obtained from commercial wood volumes is considered. These commercial volumes were derived using dendrometric parameters like base area, form factor and commercial height. Therefore, the values corresponding to monetary income represent the opportunity cost of planting a pine forest, when compared to continuing the iLUT.

Table 4.2 shows the performance attribute values for each of the 20 land unit in the dataset.

This dataset was used to study the performance of six different MCDM methods for answering the following question: “Where to afforest with pine to obtain an optimal ‘per land unit’ land performance 30 years after establishing the forest?”. The anticipated output of each MCDM method is a ranking of the 20 land units in this dataset from most to least suitable.

4.3.2 Multi-criteria decision making methods

Since a multitude of MCDM methods applicable to answer the stated question have been developed and described in literature, a preliminary exploration was conducted to select the MCDM methods to be applied. This exploration was based on seven review articles concerning multi-criteria decision analysis. Three of these articles deal specifically with forestry related problems (Kangas and Kangas (2005); Diaz-Balteiro and Romero (2008); Ananda and Herath (2009)), while the other four refer to the application of MCDM in a variety of domains (Zopounidis and Doumpos (2002)) such as sustainable energy planning (Pohekar and Ramachandran (2004); Wang et al. (2009)) and supplier evaluation and selection (Ho et al. (2010)).

Five of the considered MCDM methods, namely ELECTRE III, PROMETHEE

Land unit	Runoff mm ha^{-1} yr^{-1}	Sediment ton ha^{-1} yr^{-1}	SOC ton ha^{-1}	BOC ton ha^{-1}	Income USD ha^{-1}
1	19.50	1.59	153.64	232.66	2728.21
2	19.50	1.59	181.41	232.66	2728.21
3	30.00	1.59	153.64	232.66	2728.21
4	30.00	1.59	153.64	67.80	-129.79
5	30.00	1.59	181.41	232.66	2728.21
6	30.00	1.59	181.41	67.80	-129.79
7	30.00	2.77	153.64	163.30	1280.68
8	30.00	2.77	181.41	163.30	1280.68
9	45.24	1.59	89.63	163.30	1964.55
10	45.24	1.59	89.63	67.80	-858.56
11	45.24	1.59	89.63	163.30	678.88
12	45.24	1.59	89.63	57.02	-1951.72
13	45.24	1.59	153.64	67.80	-129.79
14	45.24	1.59	181.41	67.80	-129.79
15	45.24	1.59	329.41	67.80	-858.56
16	45.24	2.77	89.63	67.80	377.98
17	45.24	2.77	89.63	67.80	-129.79
18	45.24	2.77	153.64	163.30	1280.68
19	45.24	2.77	181.41	163.30	1280.68
20	45.24	2.77	329.41	67.80	377.98

Table 4.2: Performance attribute values for the 20 land units in the Tabacay database. Values correspond to 30 years after afforestation with pine. SOC values were computed for the first 30 cm of soil depth.

II, AHP, Compromise Programming and SMAA-2, were selected on the basis of their frequency of appearance in the aforementioned articles. ELECTRE III and PROMETHEE II belong to the family of outranking methods, AHP is an instance of a pairwise comparison method, while Compromise Programming is an example of an ideal point method, in which a decision alternative is considered better, or more suitable, when it is closer to the absolutely optimal, hypothetical alternative. The SMAA-2 method, as all other methods that belong to the SMAA family, is suitable for problems in which uncertainty is recognized in the data. IIPT was selected because it has been applied to similar problems in the past, as described in Gilliams et al. (2005b) and De Meyer et al. (2013). IIPT, like Compromise Programming, is an instance of a method based on the ideal point. IIPT iteratively defines thresholds based on the ideal point and selects at each iteration the alternatives that fulfil these thresholds (if any).

ELECTRE III

Before executing ELECTRE III (Roy (1991)), values for three parameters need to be set for each criterion, namely the preference, the indifference and the veto thresholds. To explain these thresholds, let $d_i = a_i - b_i$, where a_i and b_i represent the value that alternatives a and b, respectively, take for criterion i. That is, d_i corresponds to the difference between the values for criterion i for alternatives a and b. Then, the ELECTRE thresholds are explained as follows:

Preference threshold: If d_i exceeds the preference threshold, then alternative a is preferred over b; if not then there is no preference.

Indifference threshold: If $|d_i|$ exceeds the indifference threshold, then alternatives a and b are considered different, and vice versa.

Veto threshold: If d_i exceeds the veto threshold, a is preferred over b no matter the values of the other criteria; if not, the preference of a over b depends on the values of the other criteria.

Moreover a single value for an extra parameter, s(λ), needs to be chosen to indicate the validity of comparisons that are made to build the resulting ranking.

The different values used for the preference, indifference and veto thresholds are listed in Table 4.3. The values for these parameters were chosen by inspection of the input data (Table 4.2). In particular, the difference between the criteria values of every pair of land units was analyzed, and the parameter values were set using those differences as a basis. Considering the ranges and distribution of these differences, sensible values for two alternatives to be considered different (indifference) and for one alternative to be preferred over another (preference, veto) were chosen. For s(λ) a value of 0.15 was chosen in analogy to Raymaekers (2003).

PROMETHEE II

PROMETHEE II (Brans and Mareschal (2005)) converts differences between criterion values into a preference for one of the alternatives. To this end, a preference function has to be chosen from a group of six types: Usual, Quasi-Criterion, Linear, Level, Linear with indifference and Gaussian. These are all functions of the value of the difference between criteria values for two given alternatives. In this case, a Gaussian preference function (Equation 4.2) was selected.

$$P(d_i) = 1 - e^{-\frac{d_i^2}{2s_i^2}} \tag{4.2}$$

where $P(d_i)$ indicates the degree at which alternative a is preferred over b; d_i is the difference between values for criterion i for alternatives a and b, that is, $d_i = a_i - b_i$; and s_i is a user-defined parameter for criterion i that determines the amplitude of the curve resulting from this equation.

The Gaussian preference function was chosen because it was considered that a function that varies gradually (which is not the case for the other preference functions) represents better the 'fuzziness' regarding the preference variability of a decision maker in a realistic context. The values for s in this case were chosen as half of the maximum difference for each criterion, namely: 7.62 mm ha^{-1} yr^{-1} for runoff, 0.59 ton ha^{-1} yr^{-1} for sediment, 45.89 ton ha^{-1} 30cm^{-1} for SOC, 47.75 ton ha^{-1} for BOC and 1096 USD ha^{-1} for income.

Criterion	Preference	Indifference	Veto	Unit
Runoff	5	2	7	mm ha^{-1} yr^{-1}
Sediment	0.3	0.1	0.5	ton ha^{-1} yr^{-1}
SOC	15	5	20	ton ha^{-1} 30cm^{-1}
BOC	15	5	20	ton ha^{-1}
Income	50	20	60	USD ha^{-1}

Table 4.3: Threshold values used in ELECTRE III.

The Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP, Saaty (1977)) formulates a decision problem as a three level hierarchy. The first level corresponds to the purpose of the problem, e.g. “select the land units in which a given tree species should be planted in order to optimize certain criteria”. The second level is composed of the criteria under consideration, for example runoff or sediment production. The third level comprises the decision alternatives, in this case land units.

Once the hierarchical structure of the problem has been sketched, the decision maker, normally based on expert knowledge, must state the preferences for the second level of the hierarchy. This is carried out by means of a pairwise comparison using a scale to designate the relative importance of criteria. This scale ranges from 1 (equally important criteria) to 9 (one criterion is extremely more important than the other). Once every pair of criteria has been assigned preference value according to this scale, these preference values are used to determine a weight for each criterion.

The next step is to perform the pairwise comparison for the alternatives level. This comparison is carried out in a ‘per criterion’ basis. That is, a separate value in the scale of 1 to 9 is assigned to express how preferable is one alternative over the other one in terms of each criterion. Using these preference values, a weight is computed for every combination of alternative-criterion.

Both sets of weights, the weights corresponding to every criteria and to every alternative-criterion combination are then combined to obtain a global weight for each alternative. This weight is used to define the final ranking.

In this study, the same weight values for the considered criteria were used in all MCDM methods, therefore no pairwise comparison was necessary at the criteria level. Considering that during this study no expert knowledge was available for the pairwise comparison and preference assignment at the alternatives level, this procedure was automated using a MATLAB script that assigns preference values to every pair of alternatives based on the difference of their values for every criterion.

Compromise Programming

The first step that is performed when applying Compromise Programming (Zeleny (1973); Yu (1973)) is to compute the ideal point. The ideal point is a vector whose coordinates are given by the optimal values of the criteria, when every criterion is optimized independently of each other. The ideal point is normally infeasible, since multi-criteria decision problems involve conflicting objectives. The ideal point for the Tabacay database is shown in Table 4.4.

Solving a Compromise Programming model results in what is known as the compromise solution. It corresponds to the feasible solution (vector of criterion values for an alternative) that is closest to the ideal point. The definition of 'closeness' requires the formulation of a distance function. This is a crucial step in Compromise Programming, since the selection of a distance function will certainly determine the resulting compromise solution. An extensive explanation on Compromise Programming is given in Section 5.2.2.

Stochastic Multi-criteria Acceptability Analysis 2

Stochastic Multi-criteria Acceptability Analysis 2 (SMAA-2, Lahdelma and Salminen (2001)) is a multi-criteria decision support technique that is applied when multiple decision makers participate in the decision process. This technique is very suitable for problems with a high degree of uncertainty in the data, e.g. the criterion values. In SMAA information about the preferences of the decision makers is not necessary at all. Instead this technique determines the weight values that would make each alternative the preferred one. Additionally, this method produces indicators regarding the support for an alternative to be chosen as well as whether the accuracy of the input data is enough for making an informed decision.

SMAA-2 computes two main measures in order to assign a rank to each alternative. The first measure is called rank acceptability and it is produced for each possible combination of alternative-rank. It expresses the probability that a given alternative is assigned a certain rank. The second measure computed by SMAA-2 is called the central weight vector, which expresses, for each alternative, a combination of criteria weights that would make this alternative optimal.

Criterion	Ideal Value	Unit
Runoff	19.5	mm ha^{-1} yr^{-1}
Sediment	1.59	ton ha^{-1} yr^{-1}
SOC	329.406	ton ha^{-1}
BOC	232.662	ton ha^{-1}
Income	2728.21	USD ha^{-1}

Table 4.4: Ideal point corresponding to the Tabacay catchment. Values correspond to 30 years after afforestation with pine.

The Iterative Ideal Point Thresholding method

Gilliams et al. (2005b) introduced a MCDM method aimed at solving decision problems with a finite alternative set. They named this method Interval Goal Programming. This method was later renamed as Iterative Ideal Point Thresholding (IIPT, De Meyer et al. (2013)) to avoid confusion with a different well-established MCDM method introduced in Charnes and Collomb (1972). The workflow of IIPT is outlined below.

Step 1: In addition to the relative importance of each criterion expressed as a weight, the number of times that the algorithm is to be iterated must be specified. In this case, the number of iterations was set to 20.

Step 2: The first threshold corresponds to the ideal point. The coordinates of the ideal point hypothetical decision alternative are given by the minimal or maximal values for every criterion when considered independently of each other. The ideal point is normally infeasible, since multi-criteria decision problems involve conflicting objectives. The database is then queried to determine whether there exist land units that satisfy this threshold.

Step 3: When no land units are found that satisfy the threshold, each of the ideal point coordinates is relaxed by an interval, using Equation 4.3.

$$t_{ij} = f_i^* - j\frac{w_{max}}{w_i}\frac{f_i^* - f_{*i}}{n} \tag{4.3}$$

where t_{ij} is the value that the coordinate of the threshold corresponding to criterion i takes in iteration j, f_i^* is the coordinate of ideal point corresponding to criterion i, j is the current iteration, w_{max} is the maximum value of the weights assigned to the criteria, w_i is the weight assigned to criterion i, f_{*i} is the anti-ideal coordinate for criterion i, and n is the number of iterations.

Given the way in which thresholds are relaxed during the execution of IIPT, the whole range of values for the criterion with the maximum weight will be completely processed only after all iterations have been executed. Ranges of criteria with lower weights will be fully processed in fewer iterations. In other words, the higher the weight for a criterion, the smaller the relaxation of its corresponding threshold coordinate. This means that the exploration for criteria with higher weights is more fine-grained.

Once a new threshold has been computed by relaxing all coordinates of the previous threshold, the database is queried again to determine whether any alternatives satisfy the new threshold.

The procedure of relaxing and querying the database is repeated until the specified number of iterations is reached. The rank assigned to each alternative depends on the iteration step in which it satisfied the threshold. This ensures that, after executing all iterations, every alternative has been assigned a position in the final ranking. The algorithm does not make any provisions to avoid the assignment of the same ranking position to several alternatives.

When setting the number of iterations a sensible compromise should be made between the detail level in which the exploration is conducted and, on the other hand, the time required for IIPT to produce results. This compromise also depends on the distribution of the criteria values. In this case the number of iterations was set to 10 million.

4.3.3 Criteria weights

All the MCDM methods described above, with the exception of SMAA-2, require that the decision makers provide a weight for each criterion. These weights are values that indicate the relative importance of the different criteria (Malczewski (1999)). To this end, seventeen regional experts on land use planning in general and afforestation planning in particular who were working at the time in the study region were consulted by means of surveys. Specifically, they were asked to perform a pairwise comparison among the five criteria and assess their relative importance in the context of afforestation planning in the Tabacay catchment. This procedure resembles the steps carried out in AHP at the criteria level, in analogy to Gibney and Shang (2007). After the validation and processing the information provided by these experts, the resulting weight for each performance attribute was set as follows: 0.235 for runoff, 0.192 for sediment, 0.131 for SOC, 0.164 for BOC, and 0.278 for income.

4.3.4 Uncertainty in the database

For the purpose of assessing potential uncertainty in the input data, a partial field inventory was conducted in the study area. A comparison between the computed values and the actual measurements in the field allowed to determine a percentage with respect to the deterministic values (Table 4.2). This percentage was in turn used to define upper and lower bounds (deterministic value $\pm$ deterministic value * percentage / 100) for all criteria values for every land unit, with the aim of guaranteeing a high probability that the real attribute value is somewhere between these bounds. The percentage used for each performance attribute was: 5.5% for runoff, 13.8% for sediment, 6.6% for SOC, 38.1% for BOC, and 10% for income.

4.3.5 Software tools

Five of the six MCDM methods studied in this chapter, namely ELECTRE III, PROMETHEE II, AHP, Compromise Programming and IIPT, were implemented as scripts in MATLAB. The remaining method, SMAA-2, was executed using the free software tool JSMAA (Tervonen (2014)). The spatial output was generated using the free software geographic information software QGIS.

4.4 Results and discussion

Table 4.5 shows the ranking of the 20 land units under consideration produced by each of the six studied MCDM methods. The first column lists the numeric, sequential identifiers assigned to the land units, as indicated in Table 4.2. The numbers in the remaining columns indicate the rank assigned by each MCDM method. Figure 4.1 shows the corresponding graphical display.

Table 4.5 and Figure 4.1 show that the ranking produced by IIPT is very coarse. It distributes the 20 land units into only 6 different rank categories, which means that several land units were assigned the same position in the ranking. The most notorious case is the 9th position in the ranking, which was assigned to 11 land units. This indicates that IIPT is not able to discriminate among those land units, that is, it considers them equal in terms of their performance. It is therefore evident that in some iteration steps several land units satisfied the current threshold, while in other iteration steps no alternatives were found to satisfy it. The coarse output of IIPT greatly handicaps its comparison with the rankings produced by the other MCDM methods.

Several interesting patterns are observed when considering the results produced by all MCDM methods except IIPT. For instance, the same land units were designated as the 4 most suitable and the least suitable alternatives by all five methods. It can be said as well that, for the case of intermediate positions in the ranking, a certain degree of consistency can be observed. This applies to the ranks assigned to land units 10, 11, 12 and 16.

It is also clear from Table 4.5 that, for some land units, there is no correspondence in the ranking position assigned by the different MCDM methods. These differences might be produced by particularities both in the algorithms corresponding to each method, and in the set of parameters and their values that are used in each specific case. In general, given the different nature of each method, it is not realistic to expect that they produce identical rankings. Additionally, in the specific case of SMAA-2, the criteria data entered to the method was different from the data used by the other MCDM methods. As was explained above, interval limits for each criterion, instead of deterministic values, were used in SMAA-2. Intuitively, it can be claimed that this fact could have contributed to the inconsistencies between the ranking produced by SMAA-2 and the output of the other MCDM methods.

The coarse nature of the ranking produced by IIPT was presumably caused by the combination of two facts: the characteristics of the underlying criteria data and the particularities of the internal working of this method. It can be seen in Table 4.2 that the criteria values for all land units are distributed into few specific values. This similarity in the criteria values caused that, in some

Land unit	ELECTRE III	PROMETHEE II	AHP	CP	SMAA-2	IIPT
1	2	2	2	2	2	2
2	1	1	1	1	1	1
3	4	4	4	4	4	2
4	10	9	10	7	8	5
5	3	3	3	3	3	2
6	5	8	9	6	6	5
7	14	6	8	9	15	7
8	6	5	6	8	7	7
9	9	7	5	5	5	9
10	18	17	17	17	17	9
11	11	10	12	10	10	9
12	19	19	18	18	18	20
13	16	15	16	13	16	9
14	7	14	15	12	14	9
15	13	13	7	11	9	9
16	17	18	19	19	19	9
17	20	20	20	20	20	9
18	15	12	14	15	12	9
19	8	11	13	14	11	9
20	13	16	11	16	13	9

Table 4.5: Rankings produced by the studied MCDM methods. CP = Compromise Programming.

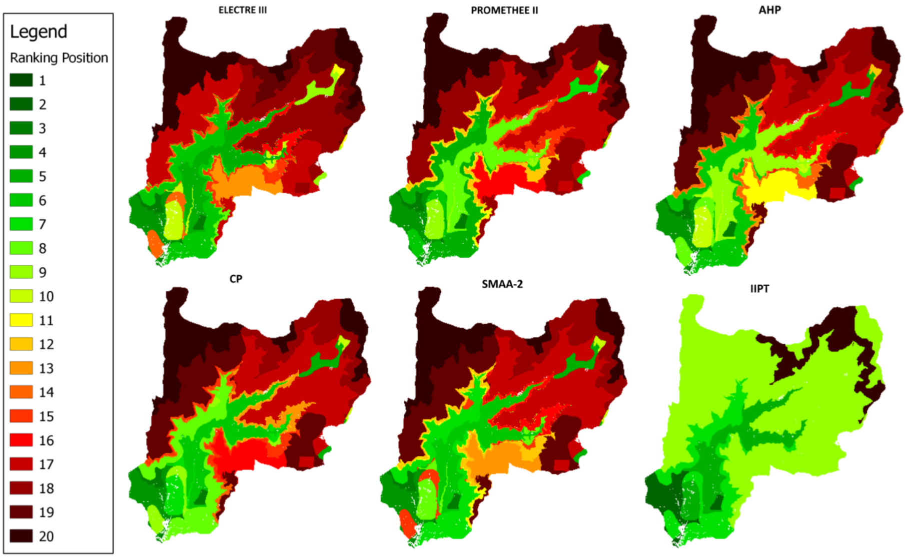

Figure 4.1: Graphical display of the rankings produced by the studied MCDM methods.

iterations, more than one land unit meet the corresponding threshold. This is the reason why, in most cases, the same ranking position is assigned to more than one alternative. Nonetheless, the most suitable alternative suggested by IIPT corresponds to the results of the other MCDM methods, as well as, at certain level, the least suitable land unit (least suitable alternative found by IIPT, i.e. land unit 12, was assigned positions 18 or 19 by the other methods). Given the relatively large number of iterations that were applied in the tests using IIPT (10 million iterations), it is not expected that using even smaller intervals (by increasing the number of iterations) would produce a more "fine grained" ranking, considering the similarity among the performance attribute values.

Alternative 17 which is designated as the least suitable alternative by all methods except IIPT presents a particular case. IIPT assigns to it the position 9 out of 20 in the ranking, which can be somewhat misleading at first sight. Nevertheless, a more careful look at Table 4.5 allows to see that the 9th position is in fact the penultimate place in the coarse ranking produced by IIPT. As a matter of fact, IIPT assigns the 9th position to 11 alternatives. Therefore we can conclude that, for alternative 17, there is also a certain degree of consistency between IIPT and the other methods.

ELECTRE III and PROMETHEE II are methods that require the specification of a relatively large number of parameters. In the case of ELECTRE III, the values for three parameters per criterion plus one global parameter must be set. PROMETHEE II requires the user to provide values for one parameter per criterion. Compromise Programming requires a single parameter to be defined. All these methods need that, in addition to the aforementioned parameters, one weight for each criterion is specified. On the other hand, SMAA-2 requires no weights and no specific parameters.

The relative consistency in the results coincides with findings reported in Gilliams et al. (2005a), who describe a comparison among six MCDM methods, namely PROMETHEE II types 1, 5 and 6, ELECTRE III, AHP and IIPT (referred to as Interval Goal Programming in that publication) when selecting tree species among three alternatives (beech, spruce and pine) to be planted on a given set of land units. However, Gilliams et al. (2005a) also report different results regarding other topics on the application of the above MCDM methods. For instance, they point out that, in many cases, ELECTRE III produces what they term as "plural solutions", i.e. solutions in which more than one alternative is recommended for a given land unit. This behaviour was not observed in the present study as shown in Table 4.5. On the other hand, Gilliams et al. (2005a) do not report this behaviour for IIPT, while in our particular case it certainly produced plural solutions. Nevertheless, the existence of plural solutions in the results of both ELECTRE III and IIPT can be considered a common situation given the characteristics of these methods. The discordance just described can be attributed to differences both in the underlying criteria data and in the tuning of the MCDM method, i.e. the specific values assigned to its parameters.

Salminen et al. (1998) reports only limited discrepancies between the rankings produced by PROMETHEE and ELECTRE III when applied to a variety of environmental decision problems. They also report the existence of plural solutions in the rankings produced by both methods. Hobbs and Meier (1994) on the other hand report more noticeable discrepancies among the results produced by the MCDM methods they compared when solving resource planning problems. Both Salminen et al. (1998) and Hobbs and Meier (1994) recommend the application of more than one method when the reliability of the results is not guaranteed in order to provide the decision maker with more insights and information about the problem at hand. Zanakis et al. (1998) report on the comparison of eight MCDM methods when applied to a large set of computer-generated problems. For these settings, they observed significant differences between ELECTRE and AHP. Summarizing, the findings reported in the literature about the comparison of MCDM methods when applied to different fields are not conclusive with respect to results consistency.

4.5 Conclusions

In this chapter a comparison among six MCDM methods for locating afforestation sites has been presented and discussed. The different methods were applied with the aim of ranking a set of 20 land units based on four environmental and one economic criteria representing the performance of the land units 30 years after afforestation with pine.

In general, a major degree of consistency was found among the rankings produced by the six studied methods. This fact allows us to claim that the selected MCDM method is not a decisive factor for the ranking, which gives the user a certain extent of freedom to select the method considering practical issues like, for example, ease of use. In the particular case of IIPT, although consistency was also observed for the most and least suitable land units, due to the distribution of the criteria values and inherent characteristics of this method, the intermediate positions in the ranking were assigned to groups of alternatives, instead of producing a fine-grained ranking as was the case for the other MCDM methods. This behaviour handicapped the comparison between the output of IIPT and the results of the other methods.

Given the very different nature of the selected methods regarding both their internal functioning as well as their parameters, there is no obvious way of guaranteeing that the selected parameter settings for each method allow for a fully consistent comparison of results. Nevertheless all methods are devised for discriminating more suitable from less suitable alternatives. The feasibility of such discrimination resides in the combination of input data and the functioning of the method at stake. When the discrimination of alternatives is feasible and the parameters are set to sensible values, the resulting ranking of alternatives will show some degree of consistency no matter the method, as was the case for most methods studied in this chapter.

Although uncertainty can be dealt with in ELECTRE III and PROMETHEE II, SMAA-2 is the MCDM method that most naturally incorporates it both into the data and into the criteria weights. On the other hand, Compromise Programming and AHP apparently leave small room for considering uncertainty.

ELECTRE III is the method that requires the setting of more parameters, while SMAA-2 does not require the specification of any parameter at all. The number of parameters required by a given method may be a relevant factor when choosing the most suitable MCDM method for solving a particular problem instance.

Chapter 5

Ideal point-based multi-criteria decision making methods for afforestation planning

This chapter is based on:
Estrella, R., Cattrysse, D., Van Orshoven, J. (2014). Comparison of three ideal point-based multi-criteria decision methods for afforestation planning. *Forests*, 5 (12), 3222-3240.

5.1 Introduction

The efforts of the Food and Agriculture Organization of the United Nations (FAO) in the 1970s resulted in the FAO framework for land evaluation (FAO (1976)) for which completed, more operational versions were published later, e.g., (FAO (2007)). In this framework, the term "land unit" is used to refer to spatially-explicit portions of land, of which the within-unit variability of diagnostic characteristics or functional qualities is smaller than the between-unit variability. This notion of "land unit" is useful in land use planning to stratify the territory of interest and as a basis for assigning LUTs to them through a matching exercise between the land unit's characteristics or qualities on the one hand, and the candidate LUTs' requirements on the other hand (FAO (1976)). Different approaches have been developed to perform this matching. One technique consists of expressing the characteristics and qualities of a land unit as a fraction of the level required by the LUT and applying the law of the minimum (Eliasson et al. (2010)). Another way is to combine the assessments of characteristics and qualities through an additive or multiplicative model, e.g., Brown (1994). As a result, maps can be produced showing for a given LUT the suitability level of each land unit. Additionally, candidate LUTs can be ranked in terms of the extent to which their biophysical and socio-economic requirements can be fulfilled by a considered land unit. Such results can be visualized in maps that show the most appropriate LUT for each land unit in the territory of interest. Both types of maps are useful to support land use planning. The former type addresses the "where" question, for example: where should a forest extension of a predefined number of hectares be established? Answering this question involves: (i) ranking the land units according to their suitability for forestry; and (ii) selecting the highest ranked land units in such a way that the cumulated area reaches the set target. The latter map type deals with the "what"-type of question: What is, among the candidate LUTs, the most suitable alternative for each land unit?

In addition to the FAO-style approaches mentioned above, several more recent MCDM methods have become available for ranking candidate LUTs according to their multi-dimensional performance on a given land unit or for ranking land units based on their multi-dimensional suitability for a given LUT. MCDM methods are specifically designed to trade-off conflicting criteria and produce a near-to-optimal solution when the absolute optimal is not achievable (Diaz-Balteiro and Romero (2008)). Such conflicts are commonly at stake in land use planning in general and afforestation planning in particular. For instance, afforestation of agricultural land typically leads to soil carbon sequestration, but also to a loss of monetary income. Some MCDM methods are applicable only to problem instances that consider a finite number of decision alternatives (e.g., AHP, ELECTRE, PROMETHEE and IIPT (4), while others are applicable to decision problems with either a finite or infinite number of alternatives (most prominently, Goal Programming and Compromise Programming). An example of a finite set of decision alternatives is the different LUTs that can be applied to a given land unit. On the other hand, in a problem that requires the optimization of the integrated land performance at a regional scale the number of alternatives is infinite. Such problems do not restrict the choice to the assignment of a single LUT to a given land unit. Instead, the decision alternatives are designated as the fractions of land units that should be covered by a LUT in order to achieve optimal regional performance. Therefore, an infinite number of LUT configurations are possible.

With the advent of the concept of ecosystem services (ESS), terminology in rural land evaluation and rural land use planning has rapidly shifted from land characteristics and qualities as defined by the FAO to the goods and services that humans experience from the (land-based) ecosystems (MEA (2005)). However, the fundamental questions have not changed: Which locations/units are most appropriate for a given LUT, i.e., "Where will the largest services or benefits be produced by that LUT?" and "What LUT will produce the largest services and benefits on a given land unit?"

In several regions of the world, land degradation due to unsustainable land use has become a major, steadily-increasing problem (Hewawasam et al. (2003); Vanacker et al. (2007); Pimentel et al. (1995)). To reverse this trend and to even improve overall land performance, the importance of the science and practice of land evaluation and land use planning cannot be underestimated (Fu and Gulinck (1994)). In this chapter, we address afforestation as a possible measure to counteract land degradation and improve land performance (Morgan (2009)). We study the question of where and with which tree species to afforest a territory of interest to achieve the best possible performance. The specific objective of this work was to test the suitability of existing methods to devise strategic afforestation plans that optimize integrated land performance at a regional scale. To this end, we applied land use allocation approaches based on Compromise Programming and Composite Programming in order to test their applicability to such decision problems, which involve the consideration of a large number of alternatives. To gain at least an initial idea about the performance of these approaches and the validity of their outcomes, we compare their results with the output of a method based on a 'per land unit' optimization. Like the approaches based on Compromise and Composite Programming, this 'per land unit' evaluation procedure uses the anticipated levels of a number of performance attributes of each land unit under each of three LUTs, i.e., continuation of the initial LUT, afforestation with pine and afforestation with eucalypt. Unlike the per-land unit approach, the goal of the Compromise and Composite Programming models is not to determine the LUT that maximize performance attribute levels for each land unit separately. Instead, they are targeted to determine how the LUTs under consideration should be distributed over the land units in order to optimize the integrated performance of the full study region.

The performance of each of these approaches is evaluated with respect to a hypothetical ideal situation, in which conflict among criteria is neglected. A first underlying hypothesis in this regard is that keeping the territory under the initial LUT distribution will result in a land performance far from this ideal, so that methods can be applied to find a LUT distribution that improves overall performance. The second hypothesis is that a LUT distribution resulting from a regionally integrated approach will produce performance levels that are closer to the ideal than the levels obtained from a ‘per land unit’ optimization.

5.2 Materials and methods

5.2.1 Study region

The study region is the catchment of the river Tabacay, described in Section 2.2.1. In this study, the Tabacay catchment was stratified into 417 land units. The definition of these land units is based on seven categorical characteristics as described in Table 5.1 (Wijffels and Van Orshoven (2009)). In Table 5.1, land characteristics are indicated as column headers, with their corresponding categories listed below them.

Each of the characteristics in Table 5.1 was available as a polygon geodataset. In spatial terms, each land unit consists of a multi-polygon resulting from the overlay of all these geodatasets. In this sense, a land unit is a (possibly scattered) multi-polygon representing an area of homogeneous characteristics. Such multi-polygons are typically different from each other in terms of shape

Initial land cover -	Soil* -	Lithology* -	Land curvature -
Crops Pasture Degraded land Eucalypt forest Native vegetation Pine forest Páramo	Suitable With restrictions Unsuitable	Suitable With restrictions Unsuitable	Convex Straight Concave
Precipitation mm yr^{-1}	Slope %	Elevation m asl	
< 650 650 - 1000 > 1000	< 25 25 - 75 > 75	< 3000 3000 - 3300 > 3300	

* Regarding suitability for forestry.

Table 5.1: Land characteristics and their corresponding classes considered to stratify the Tabacay dataset into land units.

and size. For instance the smallest land unit in this dataset has an area of 0.09 ha, while the largest one has an area of 210.6 ha. The average size of the considered land units is 15.1 ha and the standard deviation is 30 ha.

Land performance data for each of the resulting 417 land units were retrieved from Wijffels and Van Orshoven (2009). Land performance, like in Chapter 4, is expressed in terms of five on-site continuous attributes: runoff production (m^3 ha^{-1} yr^{-1}), sediment production (ton ha^{-1} yr^{-1}), stock of SOC (ton ha^{-1} $30cm^{-1}$), stock of BOC (ton ha^{-1}) and monetary income (USD ha^{-1} yr^{-1} for crops and pasture, USD ha^{-1} for pine plantation and eucalypt plantation). Runoff and sediment production are expressed as yearly rate values, while SOC and BOC represent the stock of carbon existing in soil or biomass at a given point in time.

For each combination of land unit and performance attribute, the database contains values corresponding to the land performance under the initial land cover (as derived from LandSat 7 images for the year 2002), expected performance under pine forest after 10 and 30 years, and expected performance under eucalypt forest after 10 and 30 years. The procedure used to compute each of these performance values is described in Section 4.3.1 for the case of a land unit covered by pine forest during 30 years. The same procedure was applied to compute expected land performance for all other time periods and tree species (pine forest after 10 years, and eucalypt forest under 10 and 30 years).

5.2.2 Multi-criteria decision making methods

Given that a certain level of conflict among the considered ESS was expected, which means that there does not exist a single LUT that would optimize all ESS simultaneously for a given land unit, the application of a MCDM method is a necessity. To tackle the problem at hand, three MCDM methods were chosen, namely IIPT, Compromise Programming and Composite Programming.

IIPT (Section 4.3.2) was selected considering that it has been applied to related problems in the past. In particular, Gilliams et al. (2005b); De Meyer et al. (2013); Estrella et al. (2014b) report the application of IIPT to locate sites for afforestation. There are other MCDM methods that are also suitable for performing a per-land unit optimization, as is the case for IIPT. Such MCDM methods mostly belong to the family of pairwise comparison MCDM, e.g., AHP, ELECTRE and PROMETHEE. The efficiency of pairwise comparison methods is greatly affected by the number of decision alternatives under consideration; therefore, these MCDM methods are normally applied to problems that involve only a relatively small set of alternatives. This limitation is not relevant in the case of IIPT, since it does not involve any kind of pairwise comparison. IIPT has the additional advantage of simplicity, which makes it easy to understand and implement.

Compromise Programming was selected given the feasibility of its application to problems requiring regionally-integrated optimization. The applicability of an alternative method, i.e., Goal Programming (Charnes et al. (1955); Charnes and Cooper (1957)) was also evaluated in addition to Compromise Programming. The main difference between Compromise Programming and Goal Programming is that the latter aims at satisfying a set of thresholds predefined by the decision maker, while Compromise Programming is targeted to approach the optimal solution. Although Goal Programming can be also applied as an optimizing MCDM method by setting appropriate values to its thresholds, the application of Compromise Programming, which is said to be an instance of an optimizing method, was considered more appropriate given the specific objective of achieving the best possible land performance. The same reasons are applicable to motivate the inclusion of Composite Programming in this study. In addition to the characteristics already mentioned in the case of Compromise Programming, Composite Programming has the additional advantage of favouring balanced solutions, that is, solutions for which the achievement level for all considered criteria is more or less uniform. This is not the case for Compromise Programming, in which no considerations whatsoever are made to provide for any degree of balance in its solutions. This characteristic makes Composite Programming an appealing alternative when applying MCDM methods to problems in realistic contexts, given that, in such situations, balanced solutions are normally preferable over solutions that perform well for only certain criteria and very poorly for others.

Compromise Programming

The first step that is performed when using Compromise Programming (Zeleny (1973); Yu (1973)) is to compute the ideal point. The ideal point is a vector of which the coordinates are given by the optimal values of the attributes, when every attribute is considered independently of each other. The ideal point is normally infeasible, since multi-criteria decision problems normally involve conflicting objectives.

The application of the Compromise Programming MCDM method to a concrete problem instance results in a Mathematical (typically Linear or Integer) Programming model. In case the result is a Linear Programming (LP) model, which means that both the objective function and the constraints are linear, it can be solved using the basic form of the Simplex algorithm. The Simplex algorithm cannot be applied to solve non-linear or IP models. To solve these types of models other different algorithms have been proposed. The output obtained when solving a Mathematical Programming model formulated using Compromise Programming is known as the compromise solution. It corresponds to the feasible solution that is closest to the ideal point. The definition of 'closeness' of a given solution with respect to the ideal point involves the formulation of a distance function. This is a crucial step in Compromise Programming, since the election of distance function will certainly determine the resulting compromise solution.

A first step when approaching the definition of a distance function is to determine the proximity degree between the value of the ith attribute and the corresponding coordinate of the ideal point. That is, if we represent the ideal point as in Equation 5.1.

$$f^* = (f_1^*, \ldots, f_i^*, \ldots, f_n^*) \tag{5.1}$$

where f_1^*, f_i^*, f_n^* are the optimal values for the first, ith and nth attribute, respectively, when considered independently from each other, then the proximity degree for the ith attribute can be formulated as in Equation 5.2.

$$d_i = |f_i^* - f_i(x)| \tag{5.2}$$

where $f_i(x)$ is the value of the ith attribute expressed as a function of the decision variables (x).

The next step consists of aggregating the proximity degrees of all attributes in the problem. Since attributes are typically measured in different units, they must be normalized before they can be aggregated. A typical normalization technique is given in Equation 5.3.

$$d_i = \frac{f_i^* - f_i(x)}{f_i^* - f_{*i}} \tag{5.3}$$

where f_{*i} is the anti-ideal of the ith attribute, that is the "worst" value for that attribute. After applying this normalization step, d_i will take a value in the range [0, 1]. It takes the value 0 when the attribute achieves its ideal and 1 when it is equal to its anti-ideal.

It is also important that the preferences of the decision maker are taken into account when aggregating the proximity degrees. The preferences are defined as measures of the relative importance that the decision maker assigns to each attribute. In a decision analysis context, such relative importance values are known as weights. These weights are represented by w_i in Equation 5.4, which is the general form of the distance (objective) function of Compromise Programming.

$$\text{Min } L_p = \left[\sum_{i=1}^{n} w_i^p \left(\frac{f_i^* - f_i(x)}{f_i^* - f_{*i}} \right)^p \right]^{1/p} \tag{5.4}$$

The parameter p defines the type of distance function. For example, when $p = 2$ Equation 5.4 becomes the Euclidean distance function.

For $p = \infty$ only the maximum normalized and weighted distance between any individual attribute value and its corresponding ideal is considered in the minimization process. Using D to represent this individual distance the objective function takes the form of Equation 5.5.

$$\text{Min } L_\infty = D \tag{5.5}$$

where

$$D \geq w_i \frac{f_i^* - f_i(x)}{f_i^* - f_{*i}} \tag{5.6}$$

When p is set to values other than 1 or ∞, the resulting model is non-linear, which makes the the model more difficult to solve. It has been demonstrated in the literature (Freimer and Yu (1976)) that, in decision problems involving two criteria, the points L_1 (setting $p = 1$ in Equation 5.4) and L_∞ define the bounds of the compromise set and that all other compromise solutions (solutions that correspond to $1 < p < \infty$) are situated between these bounds. For problems that involve more than two criteria, the existence of this property cannot be guaranteed. However, Blasco et al. (1999) demonstrated that under certain general conditions commonly found in (especially economic) decision problems, the property indicating that the compromise set is bounded by L_1 and L_∞ does certainly hold.

The solution corresponding to L_1 amounts to maximizing the weighted sum of the achievement of every objective. However, in some cases, this solution might be quite unbalanced, i.e. some attributes can be much closer to their ideals than others. This undesirable situation can be alleviated by using L_∞, which aims to minimize the maximum difference between an attribute value and its ideal, guaranteeing therefore a solution that is as balanced as possible.

In this chapter the p parameter was set to 1 in order to restrict the analysis to the linear case. Using Equation (5.4) as the objective function (to be minimized), replacing the parameters and adding the constraints specific to the problem at hand, the complete Compromise Programming model is expressed in Equations (5.7) –(5.8).

$$\text{Min } \sum_{k=1}^{q} w_k \frac{f_k^* - f_k(x)}{f_k^* - f_{*k}} \tag{5.7}$$

subject to:

$$x_{ij} \in 0, 1 \text{ for } i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, m \tag{5.8}$$

$$\sum_{j=1}^{m} x_{ij} = 1 \text{ for } i = 1, 2, \ldots, n \tag{5.9}$$

where:

- q is the number of criteria;
- m is the number of LUTs;
- n is the number of land units;
- x_{ij} is the fraction of land unit i that should be covered by LUT j in order to optimize the integrated land performance at a regional scale. x_{ij} are the decision variables of the problem.

A particular LUT distribution over the study region is represented in Equation 5.7 by x. A LUT distribution corresponds to a concrete assignment of values to all x_{ij}. The integrated regional performance for given attribute a corresponding to a particular LUT distribution is a function of x, and is computed using Equation (5.10).

$$f_k(x) = \sum_{i=1}^{n} \sum_{j=1}^{m} h_{ij}^k x_{ij} \tag{5.10}$$

where $f_k(x)$ represents the regional performance of a certain distribution of LUTs over the study area and h_{ij}^k is the performance value corresponding to criterion k when land unit i is covered by LUT j.

The first constraint in the Compromise Programming formulation (Equation 5.8) specifies that any decision variable can only take a value of either 0 or 1. This constraint makes this specific model an instance of an IP model. The second constraint (Equation 5.9) expresses that, from all decision variables associated to a particular land unit, only one of them can take a value of 1, while all the others must take a value of 0. In other words, this constraint imposes the restriction that one and only one LUT can be assigned to any given land unit.

The ultimate aim of any MCDM method is not to select an absolute optimal solution, since such a solution is normally not attainable given the conflict among criteria. What MCDM methods do is to generate a set of points in the solution space. This set is called the Pareto frontier (Messac et al. (2003)). One way of generating points in the Pareto frontier is by setting the parameters of the model to different values (e.g., using different sets of weight values). In this context, each point in the Pareto frontier would correspond to a particular parameter setting. The final selection of the most desirable solution among the options in the Pareto set is a subjective issue, since it is done by the decision maker, not by the method itself. The output of the MCDM-based model is just a starting point to support decisions. It is not the aim of any MCDM method to replace the decision maker. The common sense, judgement and expertise of human decision makers when making the final selection are essential in any planning process.

Composite Programming

The formulation of Compromise Programming (Equations 5.7-5.8) is focused only on minimizing the combined distance of an alternative to the ideal point. This characteristic can result in unbalanced solutions, in which alternatives that excel in some criteria but perform poorly regarding other criteria are selected as optimal, as long as their combined distance to the ideal point is smaller than other more balanced and possibly preferable alternatives. To counteract this potential lack of solution balance in Compromise Programming, the metric to be minimized in the objective function, i.e., the distance to the ideal point, must be complemented with an element that ensures a solution with some degree of balance. The use of this type of composite metrics has been already proposed in the literature with the name of Composite Programming (Bardossy and Bogardi (1983); André and Romero (2008)). The application of Composite Programming to the problem addressed in this chapter is formulated in Equations (5.11)-(5.14).

$$\text{Min} \left[\lambda D + (1-\lambda) \sum_{k=1}^{q} w_k \frac{f_k^* - f_k(x)}{f_k^* - f_{*k}} \right] \tag{5.11}$$

subject to:

$$x_{ij} \in 0, 1 \text{ for } i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, m \tag{5.12}$$

$$\sum_{j=1}^{m} x_{ij} = 1 \text{ for } i = 1, 2, \ldots, n \tag{5.13}$$

$$w_k \frac{f_k^* - f_k(x)}{f_k^* - f_{*k}} \leq D \text{ for } k = 1, 2, \ldots, q \tag{5.14}$$

There are two new elements in the objective function of the Composite Programming formulation (Equation 5.11):

- D: The balance term that corresponds to the maximum (weighted and normalized) deviation of a regional land performance level with respect to the corresponding coordinate of the ideal point. D is a (non-decision) variable of the model, for which its value is to be minimized as part of the objective function (Equation 5.11). The values that D can take are restricted by the constraint expressed in Equation 5.14, which requires the deviation from the ideal point for each and every criterion to be less than or equal to D. Given the conflict among criteria (enhancing one of them degrades some others), the minimum possible value for D will be reached when the decision variables (x_{ij}) are assigned a value in such a way that the deviations from the ideal point for all criteria are similar. This situation will occur only when full emphasis is given to a balanced solution ($\lambda = 1$).
- λ: It expresses whether emphasis is on balanced solutions (λ closer to 1) or on minimum combined distance (λ closer to 0). In other words, the objective function in the Composite Programming model becomes a linear combination between the balance term and the combined distance. Clearly, Compromise Programming is an instance of Composite Programming for which $\lambda = 0$.

IIPT and the Compromise and Composite Programming models were applied to the land units database representing the Tabacay catchment. IIPT was applied separately to each land unit in order to determine the LUT that would produce an optimal ‘per land unit’ performance while trading-off the performance attributes under consideration. “Per land unit” optimization means that the selection of a given LUT for a particular land unit is completely independent from the choice made for any other land unit. Both Compromise and Composite Programming were applied to formulate two IP models (Winston and Goldberg (2004)). The general goal of the Compromise and Composite Programming models is to optimize land performance at a regional scale. That is, in contrast to the “per land unit” approach, these models target optimization by trading-off the regionally-integrated performance of all land units with respect to the regionally-integrated ideal point.

5.2.3 Parameter settings

The combination of weights used when testing all three methods was 0.1 for runoff and SOC, 0.2 for sediment and BOC and 0.4 for income. The rationale behind the selection of this weight combination was that, usually, land owners put more emphasis on the profitability of a land use change (or continuing the current LUT), while the importance of the biophysical and environmental criteria is rated lower, but with a similar magnitude among them.

There are indeed many different perspectives besides the particular point of view of land owners. For instance, environmentalists would put more emphasis on carbon storage and soil conservation rather than monetary income, and stakeholders in the hydroelectric sector could be more interested in controlling runoff production and river sedimentation, etc. In MCDM, each of these perspectives would correspond to a particular weight setting. Furthermore, in a realistic context, it would be more sensible to run several scenarios, each of them with its particular parameter configuration, in order to better explore the solution space and allow decision makers to select the setting that seems more plausible for them, possibly considering additional requirements not expressed explicitly in the applied methods.

The number of iterations for IIPT was set to 1000. The λ parameter used in the Composite Programming model was set to 0.5, which means that equal importance was assigned to obtaining a balanced solution and to minimizing the combined distance to the optimal point. To determine an appropriate value for the λ parameter in a realistic context, a careful exploration of the underlying data would be of great help. For instance, an exploration of the data distribution for each criteria would indicate the existing similarity level among criterion values, which can give useful insights into the risk of obtaining unbalanced solutions. Additional information in this regard can be obtained from running correlation tests involving all criteria and computing the payoff matrix. Such correlation tests and the payoff matrix will provide the user also with information about the level of conflict existing among the considered criteria. This would allow, for example, removing redundant criteria from the analysis, in case a high level of correlation is detected among two or more criteria.

5.2.4 Software tools

The land units' database used in this chapter was stored using the free relational database management software PostgreSQL. IIPT was implemented in the Python programming language. The pre-processing of the input data mathematical programming models was performed using Excel spreadsheets. The conversion of the data between PostgreSQL and Excel was carried out using Python scripts. Both Compromise and Composite Programming models were implemented and executed using the solver Lingo. Their numerical output was stored in Excel spreadsheets. The link between this numerical output and the spatial land units was performed in PostgreSQL using its PostGIS extension. The final output layouts were produced in QGIS.

5.3 Results

Since all three MCDM methods are instances of ideal point methods, the first result to be computed consists of the different coordinates of such an ideal point. To compute each coordinate of the ideal point, the absolute optimal value for the corresponding performance attribute, no matter the LUT, was determined for each land unit. These performance values were then summed up for all land units to obtain a single value that represented the optimal regionally-integrated land performance. To compute the coordinates of the anti-ideal point, a similar procedure was followed with the exception that, instead of considering the optimal performance values, the "worst performance" levels (maximum runoff and sediment, and minimum SOC, BOC and income) were determined for each land unit. This procedure resulted in the values shown in Table 5.2.

The values for the ideal and anti-ideal points correspond to the regional performance computed cumulatively for a period of 30 years measured from a reference point in time in which either the land use was afforested with pine or eucalypt or the iLUT was maintained.

The results of applying the three MCDM methods to the database representing the Tabacay catchment are presented and discussed below.

The first column of Table 5.3 shows the iLUT found in the Tabacay catchment, and the last column lists the number of land units covered by these iLUT. The values in the intermediate columns indicate for each iLUT the number of land units that are suggested to be maintained under the iLUT (column labelled “Keep iLUT") or be afforested with pine (“Change to Pine”) or eucalypt (“Change to Eucalypt”). The column labelled “Keep iLUT or Change to Eucalypt” indicates that 37 land units initially under natural vegetation should either be kept like that or be transformed into eucalypt forest. This result illustrates a typical characteristic of IIPT. Since IIPT defines thresholds to be fulfilled at each iteration, there is no restriction for cases in which more than one of the decision alternatives meet the threshold at a given iteration. In such cases, IIPT will fail in establishing a distinction among those alternatives in terms of their performance and will consider them “equally optimal”, as explained in Section 4.3.2. It is also interesting to note that land units initially considered as bare land are suggested to be changed to pine or eucalypt in all cases. This seems to be reasonable, since bare lands hardly produce any income, neither do they perform well regarding the biophysical ESS. All land units under agricultural use and pasture are suggested to remain under their iLUT, mainly due the profitability of crops and livestock. Furthermore, all highlands under original vegetation (paramo) should be kept as they are according to IIPT, presumably due to their good environmental performance and despite their low profit generation. Income also plays an important role for land units initially under forest. Most land units under pine are suggested to be changed to the more profitable eucalypt, while obviously, most eucalypt forests are advised to be kept as such.

Tables 5.4 and 5.5 show the results obtained with Compromise and Composite Programming, respectively.

Attribute	Ideal point	Anti-ideal point	Unit
Runoff	215,482.93	345,157.95	10^3 m^3
Sediment	848.55	1,737.43	10^3 ton
SOC	1,207.04	597.25	10^3 ton
BOC	804.33	139.67	10^3 ton
Income	76,902.82	-610.87	10^3 USD

Table 5.2: Regional ideal and anti-ideal points used in the Compromise and Composite Programming models.

For the Compromise and Composite Programming models some trends are similar to the ones observed for IIPT: a land use change for all bare land is recommended, and most, but not all, agricultural land use is suggested to be continued. In general, land units initially under paramo are also suggested to be kept, although according to the Composite Programming model, an important share of these land units should be afforested. The trend for land units initially under forest is clearly reversed by the Compromise and Composite Programming models when compared to IIPT: the Compromise and Composite Programming models favour pine, while IIPT rather promotes eucalypt.

Tables 5.6-5.8 show a pairwise comparison of the results of the methods (IIPT vs Compromise Programming, Compromise Programming vs Composite Programming and IIPT vs Composite Programming) in the form of confusion

iLUT	Keep iLUT	Change to Pine	Change to Eucalypt	Keep iLUT or Change to Eucalypt	Total
Bare lands	0	14	27	0	41
Crops	78	0	0	0	78
Natural veg.	0	18	36	37	91
Páramo	49	0	0	0	49
Pasture	59	0	0	0	59
Pine	6	0	36	0	42
Eucalypt	44	13	0	0	57
Total	236	45	99	37	417

Table 5.3: LUT distribution resulting from the application of the IIPT method.

iLUT	Keep iLUT	Change to Pine	Change to Eucalypt	Total
Bare lands	0	41	0	41
Crops	63	13	2	78
Natural veg.	0	61	30	91
Páramo	47	2	0	49
Pasture	30	20	9	59
Pine	28	0	14	42
Eucalypt	20	37	0	57
Total	188	174	55	417

Table 5.4: LUT distribution resulting from the application of Compromise Programming.

matrices.

The first row of values in Table 5.6 corresponds to the land units that, according to IIPT, should be kept under their initial land cover (236 land units). Compromise Programming suggests keeping the initial land cover for 166 out of those 236 land units, while it recommends that 59 of them should be changed to pine and 11 to eucalypt. A similar interpretation can be done for the remaining rows. This means that the values contained in the main diagonal of the confusion matrices represent the land units for which both methods coincided in their output. As such, IIPT and Compromise Programming produced coincident outputs for 234 (sum of values in the main diagonal of matrix in Table 5.6) out of 380 land units. When IIPT assigns more than one LUT to a given land unit, the interpretation is that IIPT considers those LUTs equally good, given the intrinsic functioning of this method. Therefore, the 37 land units for which IIPT suggested more than one LUT are not considered in the analysis, since

iLUT	Keep iLUT	Change to Pine	Change to Eucalypt	Total
Bare lands	0	37	4	41
Crops	44	26	8	78
Natural veg.	0	55	36	91
Paramo	30	15	4	49
Pasture	5	35	19	59
Pine	28	0	14	42
Eucalypt	26	31	0	57
Total	133	199	85	417

Table 5.5: LUT distribution resulting from the application of Composite Programming.

		Compromise Programming			
		Keep iLUT	Change to Pine	Change to Eucalypt	Total
IIPT	Keep iLUT	166	59	11	236
	Change to Pine	0	45	0	45
	Change to Eucalypt	22	54	23	99
	Total	188	158	34	380

Table 5.6: Confusion matrix contrasting the output of IIPT and Compromise Programming. Overall agreement = 0.62.

a multi-LUT assignment, in the same sense as in IIPT, did not occur in the output of the Compromise and Composite Programming models.

Using the values contained in the main diagonal of the confusion matrices, indices for coincidence or overall agreement can be easily derived by dividing the number of land units for which agreement was observed by the total number of land units under analysis. In this way, values closer to one indicate high coincidence levels and values closer to zero, otherwise. When IIPT is contrasted to Compromise Programming, a coincidence index of 0.62 is obtained, and for IIPT vs the Composite Programming model, the index decreases to 0.5. Despite the limited number of alternative LUTs considered, this rather low coincidence index is explained by the different nature of these methods and by the differences between the per land unit approach vs regional optimization in general. On the other hand, the coincidence index reaches 0.81 for the Compromise Programming - Composite Programming comparison, which indicates a greater similarity between these methods, but it also illustrates the impact that striving for balanced solutions has on the outcome of these methods.

		Composite Programming			
		Keep iLUT	Change to Pine	Change to Eucalypt	Total
CP	Keep iLUT	127	41	20	188
	Change to pine	6	158	10	174
	Eucalypt	0	0	55	55
	Total	133	199	85	417

Table 5.7: Confusion matrix contrasting the output of Compromise Programming (CP) and Composite Programming. Overall agreement = 0.81.

		Composite Programming			
		Keep iLUT	Change to Pine	Change to Eucalypt	Total
IIPT	Keep iLUT	111	94	31	236
	Change to pine	0	45	0	45
	Eucalypt	22	44	33	99
	Total	133	183	64	380

Table 5.8: Confusion matrix contrasting the output of IIPT and Composite Programming. Overall agreement = 0.5.

Table 5.9 shows the resulting regional performance values that correspond to the LUT distribution suggested by each of the methods discussed above. The ideal point is included as a reference, and the performance that would result from continuing the iLUT on every land unit is included for comparison. Values are indicated as positive or negative deviation percentages from the ideal point.

It can be seen from Table 5.9 that the LUT distribution suggested by IIPT, when compared to continuing the iLUT, deteriorates in performance regarding runoff and sediment production, while the performance slightly and strongly improves for SOC and BOC, respectively. Regarding income, IIPT achieved the absolute optimal value. These facts are clear indicators of the stress put in IIPT to optimize the criterion with the highest relative importance to the detriment of the overall balance of the solution. When the output of the Compromise and Composite Programming models is compared to the continuation of the iLUT, it is clear that a performance improvement was achieved for all ESS, except monetary income. The decreased income performance can be explained by the trade-off that takes place in the Compromise and Composite Programming models, in such a way that the other ESS levels are enhanced (slightly in the case of Compromise Programming) at the expense of monetary income. From these deviation values, it is inferred that the Compromise and Composite Programming models are comparable in terms of their level of achievement with respect to the ideal point. These two methods surpass IIPT in this regard, except in the case of monetary income.

The spatial LUT configuration suggested by each of the studied methods is shown in Figure 5.1. The initial land use map is included for reference.

When maps (b-d) in Figure 5.1 are compared, it is noticeable that both Compromise and Composite Programming models favour the change to pine, while IIPT suggests to mostly keep land units under the iLUT or change them

	Runoff	Sediment	SOC	BOC	Income
Keep iLUT	+53	+28	-32	-81	-7
IIPT	+59	+53	-30	-57	0
Compromise Prog.	+49	+14	-23	-46	-11
Composite Prog.	+42	+22	-18	-35	-21
Ideal point	215,482.93 10^3 m^3	848.55 10^3 ton	1,207.04 10^3 ton	804.33 10^3 ton	76,902.82 10^3 USD

Table 5.9: Deviation (%) from the ideal point that would result from continuing the iLUT or implementing the LUT distribution suggested by IIPT, Compromise Programming or Composite Programming.

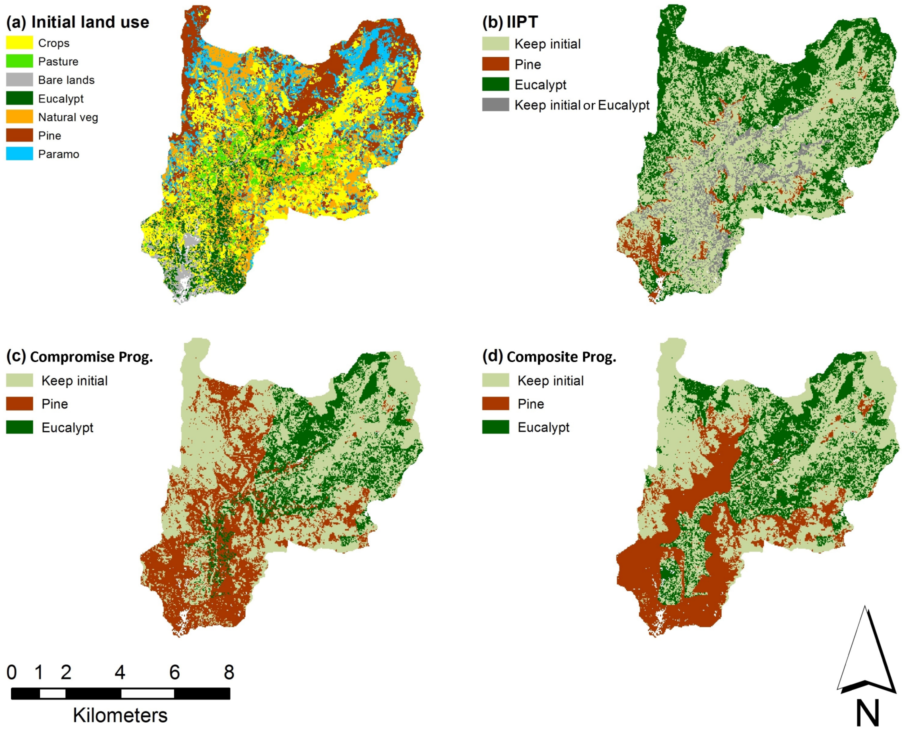

Figure 5.1: (a) Initial land use map; LUT configuration suggested by: (b) IIPT; (c) Compromise Programming; and (d) Composite Programming.

to eucalypt. It is especially remarkable that IIPT suggests to change most pine forests at higher altitudes to eucalypt. This may be an indication of the emphasis that the interval-based approach of IIPT puts on the criteria with higher relative importance (smaller intervals), as is the case for income in this study, since, according to the available data, eucalypt forests produce considerably larger profits than pine forests. In other words, IIPT focuses more on optimizing criteria with high weights at the expense of solution balance, even when compared to Compromise Programming.

The LUT distributions proposed by the Compromise and Composite Programming models are quite comparable for the particular combination of parameter values that was chosen. This similarity in the output was already revealed by the coincidence index. However, when more emphasis is given to solution balance, by increasing the λ parameter, it is expected that the difference between the output of the Compromise and Composite Programming models increases.

5.4 Discussion

Three established MCDM methods, namely IIPT, Compromise Programming and Composite Programming, were applied to a database representing the expected cumulative performance in terms of five ecosystem services of the 417 land units covering the Tabacay river catchment after 30 years of continuation of the iLUT and 30 years after afforestation with pine or eucalypt. The goal was to design a LUT configuration to be applied to the full study region in order to optimize integrated land performance.

These methods are all part of a family of MCDM methods that select alternatives based on their closeness to an ideal point, which corresponds to the optimal value of every criterion when evaluated independently of each other. IIPT was used to select the best performing LUT for every land unit separately from any other land unit. The LUT distributions generated by the Compromise and Composite Programming models, on the other hand, are targeted to the optimization of the integrated land performance of the full study region as a whole. The other difference between IIPT- and the Compromise and Composite Programming models is the way in which they search for optimal solutions. In the case of IIPT, thresholds are defined in such a way that separate deviations from the ideal value for each criterion are kept within restricted limits. In the Compromise and Composite Programming models, on the other hand, deviations from the ideal point are normalized and then combined into a single distance function, which then becomes the objective function (to be minimized) of the resulting LP models. Additionally, IIPT can only be applied to decision problems with a finite number of alternatives, while the Compromise and Composite Programming models do not have this restriction.

A particular issue in IIPT, which can be seen as a drawback in some cases, is its incapability to distinguish among decision alternatives that present a similar, although not identical, performance. Since IIPT uses an iterative procedure in which a threshold is defined and used at each step to filter alternatives with high performance, whenever more than one alternative meets a given threshold, a case of multi-alternative selection will occur. The presence of these cases in IIPT output complicates the interpretation of results and decreases the amount of information that can be distilled. Trends observed in the results indicate that IIPT is more targeted to achieve solutions that are most influenced by the criterion with the highest relative importance, which makes this method prone to producing unbalanced solutions, that is solutions that correspond to a near-to-optimal performance for the most important criterion and that perform poorly regarding the other criteria.

On the other hand, judging by the general similarity of the results produced by the Compromise and Composite Programming models, it can be concluded that they are suitable methods when a balanced solution is required. This fact is even more evident when considering the deviations from the ideal performance corresponding to the LUT distributions suggested by these methods. In particular, deviations for both methods is confined to similar levels, although this behaviour is expected to change when more emphasis is allocated to solution balance in the Composite Programming model. This expectation is not completely in line with Chang et al. (1995), who obtained virtually the same results with $\lambda = 1$ and with $\lambda = 0$ when applying Composite Programming to an environmental resources management problem, neglecting in practice the influence of the λ parameter on their model output. On the other hand, Krcmar et al. (2005) found that both instances of Composite Programming (with $\lambda = 0$ and $\lambda = 1$) produced results differing within a limited range, which is much more comparable with our findings. Clearly, discrepancies between applications of Composite Programming can be explained to a large extent by differences in the parameter settings, e.g., the relative importance assigned to criteria, and to the underlying database, in addition to the value set for the parameter λ.

For the parameter values used in our tests, Compromise and Composite Programming performed in a similar way regarding solution balance, even though this aspect is not explicitly included in the Compromise Programming model formulation. It is important to stress the importance of generalizing Compromise Programming to Composite Programming, since the explicit inclusion of solution balance considerations allows the user of this method a greater degree of control, with the capability of emphasizing either solution balance or combined optimization achievement.

Regarding modularity, since the Compromise and Composite Programming methods rely on concepts of mathematical programming, they do not impose a great deal of effort to make specific adaptations to the model formulation. In particular, to integrate restrictions with respect to minimum and maximum areas for certain LUTs or performance level thresholds to be met by the proposed LUT distribution, it would be sufficient to define and include the appropriate constraints in the model formulation. Introducing such adaptations in IIPT would undoubtedly require more effort, given its algorithmic nature and its lower level of modularity when compared to mathematical programming models.

The main restrictions for IP models, of which both Compromise and Composite Programming models are examples, reside in their limitations regarding the complexity of the problem at hand, specifically when this complexity is related to problem size. In general, the algorithms applied for solving IP models are very demanding in terms of execution time, and the feasibility of their application can degrade drastically when problem sizes exceed certain limits. In the case studied in this chapter, a limited number of land units (417) and land use management alternatives (3) were considered, which resulted in a problem instance with 1251 decision variables. This problem size proved to be quite manageable with the software and hardware infrastructure used during the experimentation phase. However, in other situations in which a more fine-grained stratification of the study region into land units is provided, and more land use alternatives are at stake, the size of the problem can restrict the applicability of methods like Compromise and Composite Programming. This problem is much less critic in IIPT, given that its algorithmic nature is rather simplistic when compared to algorithms used to solve IP models.

Whereas in this study, only on-site performance attributes like sediment production and carbon storage, were considered, a challenge is to also incorporate off-site attributes, like sediment transport and delivery (Estrella et al. (2014c); Vanegas et al. (2012)), into the optimization of the land use distributions.

Another possibility for further elaboration of the presented methods is to accommodate temporal aspects either into the algorithm, in the case of IIPT, or into the model formulation, in the case of the Compromise and Composite Programming models. The incorporation of time-related issues into the problem would require the availability of datasets corresponding to several points in time, so that answering questions like when to intervene in a territory or for how long to keep a given LUT becomes possible.

In-depth insights about the nature and functioning of these methods can be obtained from studying the impact that variations in the values of the different parameters have on the methods' output. In particular, tests involving different parameter settings would provide a clearer idea about the solution space being dealt with. In this context, a sensitivity analysis involving the λ parameter in the case of the Compromise and Composite Programming models and different weights values for all of the applied methods would shed some light on their behaviour and internal working under different scenarios and would allow the user of the methods to determine more reasonable parameter values.

5.5 Conclusion

From our comparison of these three ideal point-based multi-criteria decision methods, we recommend a regionally-integrated approach based on Composite Programming over the per-land unit IIPT approach to establish land use distributions that can serve as base maps for further operational land use planning. This suggestion does not imply that IIPT should be discarded without further consideration. IIPT can still be useful in other problem instances, as it has been shown in the past Gilliams et al. (2005b); De Meyer et al. (2013);

Estrella et al. (2014b)). IIPT can be considered as a valid alternative, especially when the input data structure permits a clear differentiation among the values of individual criteria, since this characteristic will counteract the possibility of IIPT not being able to distinguish among several alternatives. Special care when setting criteria weights is also a requirement for using IIPT. As has been shown above, differences in weights have a large impact on the method outputs. Therefore, it is advised to avoid radical differences when expressing relative importance, especially in problem instances for which reasonable improvements for all criteria involved are required.

Chapter 6

Determining land use trajectory configurations that optimize regionally integrated land performance

This chapter is the basis for:
Estrella, R., Cattrysse, D., Van Orshoven, J. (under review). A Composite Programming model to determine land use trajectories for optimizing regionally integrated ecosystem services delivery. *Forests: Special Issue Ecosystem Services from Forests.*

6.1 Introduction

Land use planning is an essential tool for achieving the ultimate goal of sustainable territorial development (Godschalk (2004)). Land use planning requires making decisions that involve multiple environmental and socio-economic criteria (Joerin et al. (2001)). Such decisions are typically based on two approaches. On the one hand, they are related to the suitability of a land unit (portion of land with uniform characteristics (FAO (1976), FAO (2007))) for establishing a particular LUT; and, on the other hand, they are targeted at determining the LUT that would optimize the performance of a specific land unit. The former is an instance of a site location problem: Where should a given LUT be established?, while the latter corresponds to solving a *What?* question: What LUT should be applied to a given area? Due to the fact that land units are not isolated entities, but interacting components of a larger system, i.e., the region of interest, independent land performance assessment of separate land units is of limited usefulness. Instead, enhancing regional performance, expressed as the integral contribution of all individual land units, is typically the goal of most land use planning projects.

The utility of land use planning becomes apparent when considering its goals and consequences in the long term, i.e., land use planning objectives cannot be confined to the immediate future. This is particularly true when land use planning involves LUTs that produce gradual effects on the environment, such as forests. In such cases the incorporation of the temporal dimension in the problem becomes a requirement. This leads to a more elaborated version of the *What?* question, in the sense that land use management of a region over a medium or long term will most certainly require the definition of LUT sequences that should be applied within a time span in order to enhance regional land performance. The temporal dimension in this context can be considered in two ways: absolute, when land performance is assessed with reference to a particular period in time, e.g., the period 2015-2045; and relative, when land performance is not linked to any specific point in time, i.e., land performance in this case is assessed for a reference, abstract, period of time. Considering time in the absolute way can be claimed to be more realistic but it is also more complex, since in that case the expected land performance corresponding to a specific point in time would need to be determined. To compute this performance all particular conditions existing at that particular point in time, e.g. climate, should be determined and involved in the estimations. Not linking the study period to any specific point in time offers a more abstract picture of reduced complexity.

In Chapter 5 a Mathematical Programming model was formulated and applied to determine the land use configuration that would optimize regional land performance aggregated over a 30 year period without considering land use trajectories. This model was formulated using the Composite Programming MCDM method (André and Romero (2008)). The term ‘land use configuration’ is used in Chapter 5 to refer to a specific distribution of a predefined number of LUTs over the study region with a view to fulfil multiple criteria. Besides formulating and applying this model, Chapter 5 reports on the comparison of its output with respect to other MCDM methods.

In the present study we build further upon Chapter 5 to address the optimization of the land performance of a region considering the potential implementation of all possible combinations of four LUTs: crops, pasture, pine plantation and eucalypt plantation. The Mathematical Programming model formulated in Chapter 5 was used to determine the way in which these LUTs should be sequenced over the land units in the study area and over a time span of 30 years, with the aim of achieving optimal cumulative regional performance. Like in Chapter 5, land performance was expressed as the trade-off of five continuous and conflicting attributes: runoff production, sediment production, organic carbon stock in soil, organic carbon stock in biomass, and monetary income. Given that the study region is represented as a set of land units, the output of this model is a configuration in which each land unit is assigned a LUT sequence, or trajectory. Regarding the temporal dimension, 10 year intervals covering a time span of 30 years are considered. In addition to test the performance of this model for the data initially available and with a set of fixed parameter values, in this chapter its behaviour is also assessed under conditions of uncertainty in the input data and its sensitivity to varying parameter values is evaluated. Moreover, the way in which the results are affected when including performance thresholds as additional constraints is investigated.

6.2 Literature review

Land use planning and management, both in urban and rural areas, can be tackled using a multitude of different approaches. Such approaches are typically members of one of two broad families of methods: (i) exact techniques (e.g., mathematical programming models), or (ii) (meta-)heuristic methods. Principles taken from these two areas have been applied to allocate different uses to specific area units within a region. In this context, these techniques must take into account the fact that land use planners are typically interested in several criteria of different nature. Furthermore, some degree of conflict is commonly observed among these criteria, in the sense that enhancing one of them decreases the level of one or more of the others. In such cases, the application of MCDM methods (Belton and Stewart (2002), Gal et al. (2013)) becomes a necessity. These methods are well suited to find near to optimal solutions while trading off the conflicting criteria at stake. They also typically allow to express differences on the relative importance of these criteria.

Besides the method formulated in Chapter 5, which is further elaborated in the present chapter, several other examples of multi-criteria mathematical programming models exist. For instance, Chang and Ko (2014) introduces a dynamic mathematical programming model that is capable of suggesting optimal land use plans. This model considers several environmental (carbon emissions in particular and environmental pollution in general) as well as socio-economic criteria (economic development and employment opportunities).

It also integrates the decision maker preferences in the form of a so-called compromise index. Another example of an MCDM for land use planning is Chen et al. (2011), which describes the application of the Ordered Weighted Averaging multi-criteria evaluation method in order to assess the safety of sites based on several criteria like earthquake risk and site type. The latter is an example of an exact method that does not pertain to mathematical programming. Darradi et al. (2012) applies the Goal Programming MCDM (Charnes and Cooper (1957)) to optimize the environmental performance of agriculture. This application starts by defining a system of 'high environmental state' as the goal at which the current environmental state is aimed. This model generates land use plans based on the optimization of three environmental criteria: nitrogen in water, sediment in water and water yield. Additionally, Vanegas et al. (2012) compares the performance of an IP model and a heuristic technique when locating areas within a river catchment that should be afforested in order to minimize sediment yield.

Besides Vanegas et al. (2012), several other examples of heuristic and meta-heuristic methods can be cited. For instance, Cao et al. (2012) introduces a method based on a genetic algorithm to determine land use plans that optimize socio-economic, environmental, and even topological (compactness) criteria. Another example of an application of a genetic algorithm in land use planning is reported in Holzkämper and Seppelt (2007), which describes a genetic-algorithm-based software library that can be applied to optimize land use configurations. It uses a patch topology representation to stratify the study region represented as a raster into areas of contiguous cells. Each of these areas, or patches, is considered as an individual entity, to which a specific LUT can be assigned. Another novel representation of geographical entities is introduced in Xiao (2008), which describes the use of graphs to represent a geographical region. This representation is then used as the input for evolutionary algorithms, the general class of which genetic algorithms are instances, to solve geographical optimization problems, like area selection or region partitioning. Duh and Brown (2007) uses the principles of simulated annealing in order to develop a method for determining optimal land allocation patterns that fulfil multiple objectives.

Mathematical programming and (meta-)heuristics are not mutually exclusive. Techniques that combine principles from both areas have also been reported in the literature. An example is Aerts et al. (2005), which formulates a Goal Programming model to solve a multi-site land use allocation problem. This type of problems consists in assigning more than one LUT to a given area. Since the resulting Goal Programming model is non-linear, meta-heuristics are applied to solve the problem in a reasonable time. In particular, the performance of a genetic algorithm is contrasted to that of a simulated annealing approach.

Topological factors like compactness and contiguity are considered among the optimization criteria.

Several studies about sensitivity analyses of multi-criteria land use planning methods and their behaviour under data uncertainty have been conducted in the past and reported in the literature. For instance, Chen et al. (2010) describes the analysis of the sensitivity of MCDM in the context of a Geographic Information System to variations on the relative importance values of each criterion. The simulations performed in Chen et al. (2010) also allowed to determine the criteria that were particularly sensitive to such variations. In Tenerelli and Carver (2012) a method for determining the suitability of sites for conversion to perennial bioenergy crops is introduced. A simulation approach was applied to evaluate the influence of uncertainty of the input data and parameters on the model behaviour and outcomes. A closely related study is reported in Verstegen et al. (2012), which introduces a model that integrates simulation, and analysis and visualization of uncertainty in the context of spatial decision support. The presented tool includes a spatio-temporal modelling framework, which simulates land use changes dynamically over several time steps. It also incorporates a Monte Carlo analysis framework that, based on the simulation component, is capable of producing stochastic maps. An assessment of the impact that uncertainty in both the input variables and the parameter settings have on the outcomes of this model is discussed.

6.3 Materials and methods

6.3.1 Study region and available data

The study region is the Tabacay river catchment, described in Section 2.2.1. The major geodatabase used in this chapter is similar to the one used in Chapter 5. In particular, the same five on-site performance attributes are considered. For each land unit, information about these attributes is available for its initial land cover (PROMAS (2005)), and for all other LUTs. This database also contains estimated performance attribute values for pine or forest plantations of 10 and 30 years of age, for each land unit. The values corresponding to runoff production, sediment production and SOC, at years 0, 10, 20 and 30 for crops and pasture are assumed to remain at the same level over time. For these two LUTs BOC is assumed to always be 0 and monetary income at each 10 year period corresponds to the sum of the yearly values.

The land units covered by páramo as well as those covered by natural woody vegetation were excluded from the study, since it was considered that it is

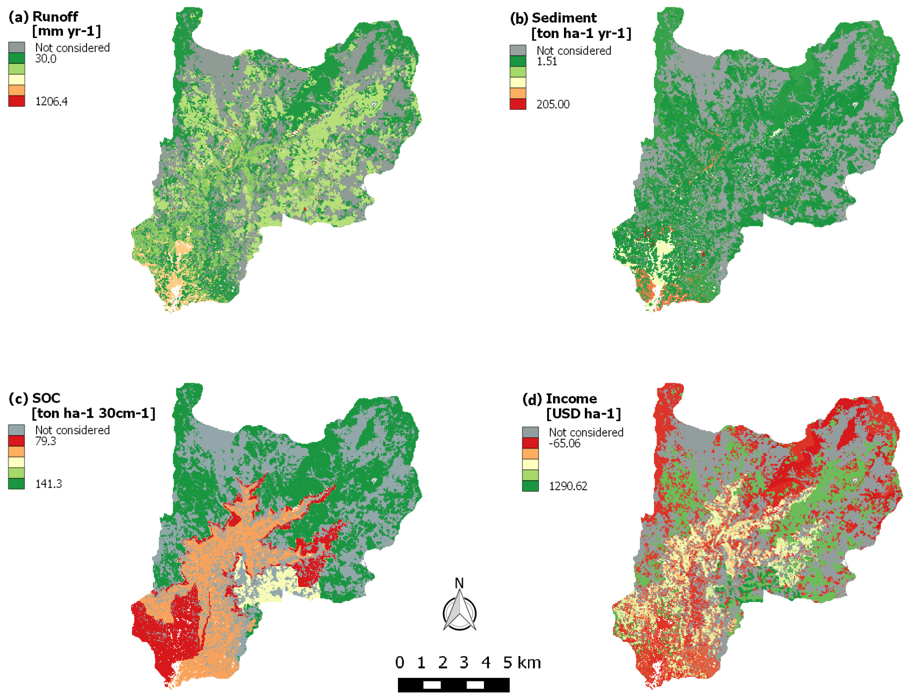

Figure 6.1: Initial land performance of the Tabacay catchment expressed in terms of (a) Runoff production, (b) Sediment production, (c) SOC, and (e) Monetary income. The areas in gray correspond to land units under páramo and natural vegetation, which were excluded from the study.

ecologically important that such LUTs remain undisturbed. This filtered out 140 land units from the database, limiting the analysis to 277 land units (66%) that represent 41.83 km^2 (62.9%) of the Tabacay catchment area. Each of the remaining land units can initially be under one of five possible land covers (LUT before year 0): crops, pasture, degraded land, pine forest and eucalypt forest.

The maps in Figure 6.1 show the land performance of the Tabacay catchment under the initial land cover, as expressed by its performance attribute levels. BOC is not included in Figure 6.1 because it is assumed to be 0 for the full catchment under the initial land cover.

6.3.2 Land use trajectories and their performance

The present study aims at determining, for each land unit, the land use trajectory that contributes to the optimal cumulative land performance of a region over 30 years. The term land use trajectory refers to a series of LUTs that are implemented sequentially within a time span of 30 years at 10 year intervals. Four LUTs are considered: crops, pasture, pine plantation, and eucalypt plantation.

A land use trajectory for a land unit is defined as follows. First, one of the four considered LUTs is assumed to be established at the initial year (year 0), and this LUT is kept in place until year 10. At this point the current LUT can be kept as such or be replaced by a different LUT. The LUT established at year 10 is again assumed to be kept until year 20, when it can be kept or changed to another LUT. Finally, the LUT implemented at year 20 is kept for the remaining 10 year period. Therefore, a land use trajectory is defined by an ordered sequence of three LUTs that are to be implemented at 10 year intervals, starting at a certain reference year (year 0, relative time) and within the boundaries of a 30 year time span. Both the 30 year time span and the 10 year intervals at which a land use change may or may not occur were chosen considering the available data. An alternative, more flexible approach would have been to determine the specific point in time at which land use is changed as part of the decision process. In this case in addition to selecting the target LUT to apply, the specific year at which such land use change should occur would be determined. This approach is clearly much less restrictive than the definition of land use trajectory used in this chapter, however it would add more complexity to the model formulation and it would drastically increase the input data and computational requirements.

Any LUT present in a given land unit before year 0 is assumed to be replaced by the first LUT of the sequence corresponding to the land use trajectory. For instance, if the LUT before year 0 (or initial LUT, iLUT) is pine forest and the first LUT in a given land use trajectory is also pine forest, the existing forest is assumed to be cut and replanted instead of being continued. The same reasoning applies for the case of eucalypt forest. There are two possibilities at years 10 or 20 regarding previously existing forest: either the forest is kept in place and continued for the next time interval, or the trees are cut and replaced by crops, pasture, forest of a different tree species, or by a new forest of the same tree species. For instance, for a given land use trajectory that involves the implementation of eucalypt forest from year 0 to year 10, such forest can be continued (so that a 20 years old forest will be present at year 20), cut and replanted, or replaced by crops, pasture or pine forest. Each of these possibilities corresponds to a different land use trajectory. At year 30 any existing forest is assumed to be harvested. Likewise, crops and pasture can be either continued

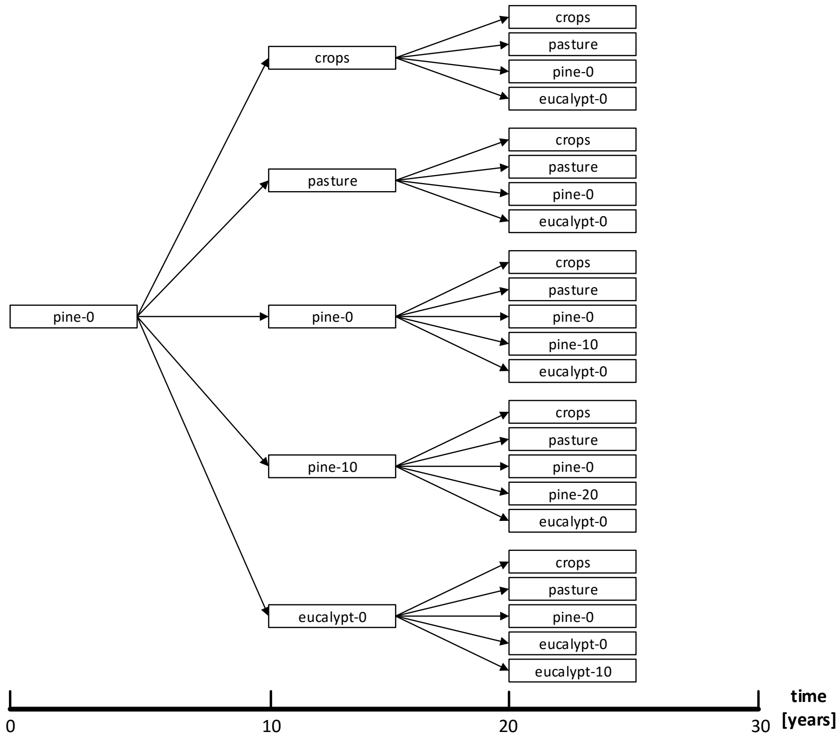

Figure 6.2: Tree view of all land use trajectories that start with pine forest. pine-X or eucalypt-X indicate a pine or eucalypt forest of X years of age, respectively.

or replaced by a different LUT at years 10 or 20.

Following the principles sketched above, each possible sequence of the considered LUTs defines a separate land use trajectory, which results in a total of 82 possible trajectories. A tree view of the 23 land use trajectories that start with pine forest is shown in Figure 6.2.

The performance level of a particular land unit at a given point in time depends on the current and, possibly, on previous LUTs that have been established on it. Performance values were determined for each possible combination of the 277 land units and 82 land use trajectories. For each of these combinations, a performance value for each of the considered points in time (years 0, 10, 20 and 30) was determined. This procedure was applied to each of the five performance attributes under analysis. All values that were initially not available (values for years 10, 20 and 30 under crops and pasture and values for year 20 under pine and eucalypt) were computed from the information that was available in the database described in Section 6.3.1 following the assumptions summarized below.

1. Land performance was assumed to remain constant under crops and pasture, i.e., performance values for years 0-10, 11-20 and 21-30 for crops and pasture were assumed to be the same;

2. The land performance for a land unit under a 20 years old pine or eucalypt forest was linearly interpolated using the values corresponding to forests of 10 and 30 years of age (which were available in the database).

To determine the performance at year 0 of a land unit under a LUT other than its iLUT, a procedure based on similarity among land units was applied. Land unit similarity was defined in terms of land characteristics. Two land units are maximally similar when six (out of seven) of their diagnostic land characteristics coincide, and minimally similar when they do not coincide in any of their land characteristics. Based on this land similarity principle, the procedure below is then applied. In this explanation, the term current land unit is used to refer to the land unit for which performance at year 0 is being computed, and the term target LUT is used to refer to the LUT different than the current land unit's iLUT for which the corresponding performance level needs to be determined.

The first step consists in searching the database for land units that are both maximally similar to the current land unit and for which their iLUT corresponds to the target LUT. Three different outcomes can result from this search:

1. Only a single land unit is found to meet the requirements stated above: in this case, the performance value corresponding to the iLUT of this land unit (target LUT) is assigned to the current land unit;

2. More than one land unit is found to meet the requirements: the average of the performance values corresponding to the found land units is assigned to the current land unit;

3. No land units are found to meet the requirements: the similarity requirement is relaxed by removing one land characteristic (e.g., instead of requiring a coincidence in n land characteristics, the requirement is restricted to n - 1 characteristics). The land characteristic that is removed from consideration depends on the performance attribute for which its value is being computed. In principle, this relaxation (generalization)

increases the probability of finding land units that fulfil the requirements in the next iteration.

This “search and generalization” procedure is iterated until a match with one or more similar land units under the target LUT are found. This procedure is applied to all land units under consideration, until every land unit has been assigned a performance value at year 0 for all attributes (runoff, sediment, SOC, BOC and income).

Finally, the effects on performance of LUT transitions at year 0, 10, and 20 needed to be determined. To this purpose, additional assumptions depending specifically on the performance attribute being considered were made. Using all the general assumptions explained above in combination with the specific assumptions explained below for every performance attribute, a curve corresponding to every possible combination of land unit (277), performance attribute (5) and land use trajectory (82) was computed. Such curves are defined by the level that a given performance attribute reaches at every point in time (year 0, 10, 20, and 30) when a decision on maintaining or modifying the LUT in a specific land unit is made.

Based on these assumptions, performance values for years 0, 10, 20 and 30 were computed for each possible combination of land unit, land use trajectory and performance attribute. The results of the computation of these performance values are explained below for a sample land unit and trajectory.

The sample land unit’s characteristics are is presented below.

- Area: 2.7 ha
- Initial land cover: Pasture
- Soil: Suitable with restrictions for forestry
- Lithology: Suitable for forestry
- Slope: 25-75%
- Land curvature: Convex
- Elevation: 0-3000 m asl

Table 6.1 shows some statistical indicators aggregating all the trajectories corresponding to the sample land unit.

The sample land use trajectory involves replacing the initial pasture by pine forest (pine-0) at year 0, then replacing the 10 years old pine forest (pine-10) by eucalypt forest (eucalypt-0) at year 10, and then replacing the 10 years old eucalypt forest (eucalypt-10) by pasture at year 20, and keeping pasture between year 20 and 30. In the sections below, the specific assumptions to compute performance values in the case of each attribute are explained. Illustrations are provided using the sample land unit and trajectory.

Runoff production

Besides considering the general assumptions explained above, the computation of performance values for runoff was based on the following specific assumptions:

1. If at year a given land unit is afforested, the runoff at year 0 is assumed to be at the level corresponding to the LUT present before year 0 (either degraded land, crops or pasture);
2. If at year 0 a forest is cut and replanted, the runoff level is assumed to be the one corresponding to crops. This principle is based on the assumption that when forest is cut the land is cleared, and when forest is replanted the soil is perturbed, and that these interventions take the soil to a runoff state similar to when it is under crops;
3. At any year (0, 10, or 20), an instantaneous change in runoff is assumed when the LUT is changed to crops or pasture. For instance, if at year 0 a land unit corresponding to degraded land is covered by crops, it is assumed that the runoff level immediately reaches 'crop performance';
4. If at year 10 or 20 forest is cut and replanted (either pine or eucalypt), runoff is assumed to return to a level corresponding to 90% of the performance level that the land unit had before afforestation, in case forest was present on the land unit for no more than a 10 years period.

Attribute	Unit	Min	Max	Avg.	St. dev.
Runoff	$m^3\ ha^{-1}$	49,500	310,000	134,660	68,153
Sediment	$ton\ ha^{-1}$	187	6,150	3,024	1,450
SOC	$ton\ ha^{-1}\ 30cm^{-1}$	94	387	175	114
BOC	$ton\ ha^{-1}$	0	407	141	86
Income	$USD\ ha^{-1}$	2,281	42,708	13,809	9,341

Table 6.1: Statistical indicators for the cumulative performance values over 30 years of all considered land use trajectories corresponding to the sample land unit.

This assumption is based on the fact that some vegetation develops under the tree canopy while forest is present in a given land unit. This vegetation is assumed to contribute to decrease the runoff level, also after trees are cut. When forest has been present for more than 20 years on a land unit and then it is cut and replanted, runoff is assumed to reach a level of 81% with respect to the land unit performance before afforestation, since vegetation underneath the trees' canopy has had a longer time to develop, and therefore it contributes more importantly to the decrease of runoff;

5. When a 10 year forest interval (e.g., eucalypt from year 10 to 20) follows a 10 year interval also under forest (e.g., pine from year 0 to 10), the runoff level reached at year 20 corresponds to 90% of the performance of a 10 years old eucalypt forest. This assumption considers the fact that some vegetation developed during the previous time interval (pine between year 0 and 10, in this example) and that it contributes to the decrease of the runoff level. When a 10 year forest interval (e.g., pine from year 20 to 30) follows two 10 year intervals under forest (e.g., pine from year 0 to 10 and eucalypt from year 10 to 20), the performance level at year 30 is assumed to be 81% of the performance corresponding to a 10 years old forest, since vegetation has had a longer time to develop underneath the canopy and, therefore, it contributes to a larger decrease in runoff;

6. If at any year (0, 10 or 20) the current LUT is maintained, no immediate change in the runoff level is assumed to occur.

Table 6.2 lists the runoff production values computed using both the general and the specific assumptions for the sample land unit. These values correspond to the breakpoints of the time-performance curve for the sample land use trajectory shown in the graph to below the values.

Although according to the sample land use trajectory at year 0 the land unit is afforested with pine, the value of runoff production still corresponds to pasture, given that, based on the assumptions above, establishing a forest is assumed not to produce immediate effects in the runoff level. Runoff is assumed to evolve linearly until it reaches the known level corresponding to a 10 years old pine forest at year 10. At this point the pine forest is replaced by a eucalypt forest, which causes runoff to immediately go back to 90% of its original level under pasture (since some vegetation is assumed to have developed under the trees during the period in which the pine forest was present, and such vegetation contributes to runoff reduction). Then runoff evolves until it reaches 90% of the known runoff corresponding to a 10 years old eucalypt forest. Similarly, the factor of 90% is used a second time to take into account the vegetation that developed underneath the trees during the two periods. At year 20, the

10 years old eucalypt forest is converted to pasture, which causes runoff to immediately reach the level of the new LUT. Between year 20 and 30 the runoff level under pasture is assumed to remain constant. Since runoff production is a rate variable, which means that it expresses the amount of runoff that is produced on a yearly basis, the total cumulative runoff production over 30 years corresponds to the shaded area under the curve in the graph.

Sediment production

The specific assumptions made to compute sediment performance values are exactly the same as the ones made for the case of runoff, which are detailed in the numbered list at the beginning of Section 6.3.2. Table 6.3 shows the sediment breakpoint values for the sample land use trajectory and land unit and the corresponding graph.

The sediment value at year 0 corresponds to the iLUT, i.e., pasture, since the change to pine does not produce an immediate impact on the land unit's

Time period	LUT	Runoff production $m^3\ ha^{-1}\ yr^{-1}$
iLUT	pasture	3000
from 0	pine-0	3000
until 10	pine-10	2100
from 10	eucalypt-0	2700
until 20	eucalypt-10	1950
from 20	pasture	3000
until 30	pasture	3000

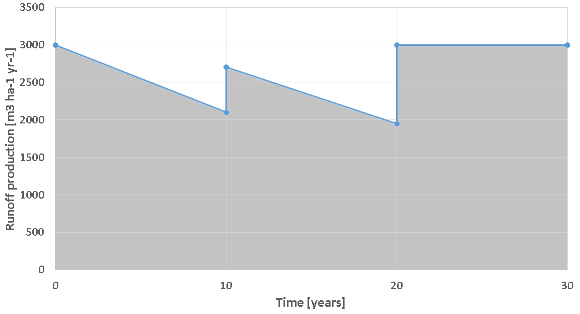

Table 6.2: Runoff production values and graph for sample land unit and land use trajectory. Cumulative runoff production over 30 years: 78,750 $m^3\ ha^{-1}$.

performance. At year 10 sediment reaches the level corresponding to a 10 years old pine forest. This pine forest is then cut, causing sediment to go up to 90% of the original level under pasture. The pine forest is replaced by a eucalypt forest at year 10, which causes that sediment at year 20 reaches a level corresponding to 90% of sediment production under a 10 years old eucalypt forest. At this point this eucalypt forest is harvested, thus the soil is disturbed and pasture is established, which makes sediment to immediately reach the level corresponding to pasture and remain at that level for the following 10 years until the end of the 30 year time span.

Stock of Soil Organic Carbon (SOC)

Since in the input database SOC values represent carbon stocks available in the soil at any given point in time, no immediate changes can occur when a transition is made from one LUT to another. A linear evolution is assumed within every 10 year interval. For example, if at year 0 a pine forest is established

Time period	LUT	Sediment production ton ha^{-1} yr^{-1}
iLUT	pasture	205
from 0	pine-0	205
until 10	pine-10	40.56
from 10	eucalypt-0	184.52
until 20	eucalypt-10	61.63
from 20	pasture	205
until 30	pasture	205

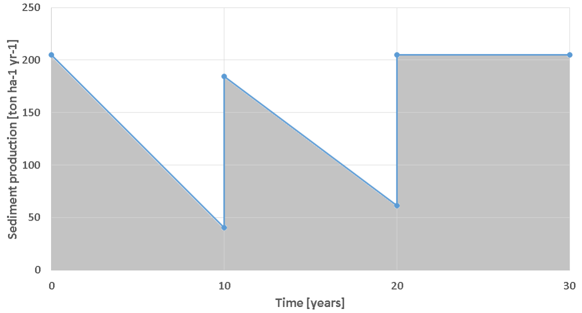

Table 6.3: Sediment production values and graph for sample land unit and land use trajectory. Cumulative sediment production over 30 years: 4509 ton ha^{-1}.

in a given land unit, the SOC level at that point is assumed to correspond to crops (cleared land, disturbed soil). SOC is then assumed to linearly evolve throughout 10 years, until it reaches the stock level corresponding to a 10 years old pine forest. This reasoning is applied to every 10 year interval until the stock of carbon in the soil at year 30 is determined.

The SOC breakpoint values and the corresponding graph for the sample land unit and land use trajectory are shown in Table 6.4.

Unlike for runoff and sediment, SOC levels do not change immediately, no matter the land use change applied. Although according to the sample land use trajectory the iLUT is changed to pine forest at year 0, the SOC level still corresponds to pasture. Then SOC is assumed to linearly evolve from pasture level at year 0 until the level of a 10 years old pine forest is reached at year 10. At this point the pine forest is replaced by a eucalypt forest, causing SOC to evolve linearly until the level corresponding to a 10 years old eucalypt plantation at year 20. At year 20 eucalypt is replaced by pasture, therefore the SOC level evolves until it reaches a value corresponding to pasture at year

Time period	LUT	SOC ton ha^{-1} 30cm^{-1}
iLUT	pasture	93.56
from 0	pine-0	93.56
until 10	pine-10	105.81
from 10	eucalypt-0	105.81
until 20	eucalypt-10	122.96
from 20	pasture	122.96
until 30	pasture	93.56

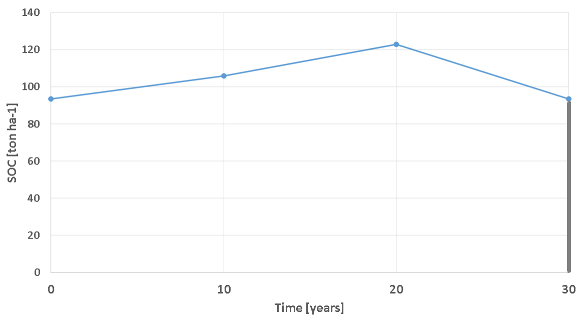

Table 6.4: SOC stocks and graph for sample land unit and land use trajectory. Total SOC stock after 30 years: 93.56 ton ha^{-1}.

30. Since SOC is a stock variable, which means that SOC values represent the amount of carbon accumulated in the soil at a particular point in time, the final aggregate value considered to represent SOC performance for every land unit - land use trajectory combination is the level reached at year 30, which is indicated by the bold vertical bar in the graph.

Stock of Biomass Organic Carbon (BOC)

BOC values in the input database express the amount of carbon stored in wood. Based on this fact, values of BOC at year 0 were assumed to be 0 in all cases. At any other points in time (year 10, 20 or 30) a BOC value was considered only when a forest was harvested. This assumption implies that for crops and pasture, as well as when either a pine or eucalypt forest are maintained, BOC values are not considered (BOC = 0). The value for BOC at year 0 is always considered to be 0, even when a pre-existing forest is assumed to be cut. This assumption follows from the fact that BOC levels depend on the forest's age, which is unknown for forests existing before year 0. The consideration of these assumptions resulted in the values listed and graphed in Table 6.5.

It can be seen in the graph in Table 6.5 that, following the assumption stated above, the level of BOC at year 0 is set to 0. Since a pine forest is planted at year 0 and harvested at year 10, BOC is assumed to linearly evolve until the known level for a 10 years old pine forest. The pine forest is cut at year 10, causing the BOC level to drop immediately to 0. Then a eucalypt forest is established, and BOC evolves until the level corresponding to a 10 years old eucalypt forest, which is harvested at year 20. This makes the BOC level to drop instantaneously to 0, where it remains until year 30, since the BOC level for pasture is assumed to always be 0. The BOC levels for the points in time at which a forest is harvested are simply added up to compute the final aggregate value for the full land use trajectory. In the example shown in Table 6.5, BOC levels at years 10 (116.22 ton ha^{-1}) and 20 (55.78 ton ha^{-1}) are summed up to obtain the final BOC aggregate value (172 ton ha^{-1}) for the sample land unit and land use trajectory.

Monetary income

Some of the assumptions made to compute income values are similar to the ones used for BOC. Like for BOC, the income level at year 0 is always considered to be 0, since income depends on the period of time a given LUT has been present on the land unit, which is not known in this case. On the other hand, unlike the case of BOC, income produced by non-forestry LUTs are certainly taken into account. Regarding these LUTs, income was computed at every point in time (year 10, 20 and 30) besides year 0. Since the available information about income produced by non-forestry LUTs indicates yearly values, these values have to be integrated over the period of time for which crops or pasture have been present on a land unit, in order to compute the cumulative income. For pine or eucalypt forests, on the other hand, income values are only considered when wood is harvested from the forest, as for the case of BOC, therefore income was assumed to be 0 at any point in time when a forest is continued. Another difference between BOC and income is that it is possible to find negative values in the case of income. Negative income values represent the situation in which the current LUT (e.g., pine or eucalypt forest) has generated more costs (planting, harvesting, transport, etc.) than revenue. The income values and the corresponding graph for the sample land unit and land use trajectory are shown in Table 6.6.

In Table 6.6 the income level at year 0 is assumed to be 0. At year 10, the income produced by taking all the wood from a 10 years old pine forest is considered. A similar situation occurs at year 20, when a 10 year old eucalypt

Time period	LUT	BOC ton ha^{-1}
iLUT	pasture	0
from 0	pine-0	0
until 10	pine-10	116.22
from 10	eucalypt-0	0
until 20	eucalypt-10	55.78
from 20	pasture	0
until 30	pasture	0

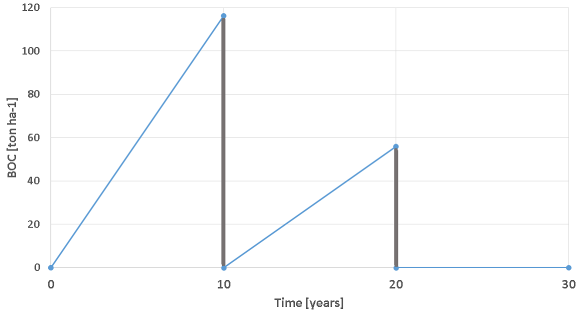

Table 6.5: BOC stocks and graph for sample land unit and land use trajectory. Cumulative BOC taken from forest over 30 years: 172 ton ha^{-1}.

forest is harvested. Between years 20 and 30, the yearly income produced by pasture is integrated over the 10 year period, as indicated by the shaded area. To compute the total cumulative income over the 30 years time span, the values obtained when a forest is harvested (vertical bars at years 10 and 20) are added to the total income produced by pasture between year 20 and 30 (shaded area).

6.3.3 A Composite Programming model for land use planning with a view to optimizing regional land performance

In Chapter 5 (Section 5.2.2) a model based on the Composite Programming MCDM method was formulated and applied to determine the land use configuration that would optimize regional land performance aggregated over a 30 year period without considering possible land use trajectories over time. The same model was used in the present chapter to find the land use trajectory

Time period	LUT	Income USD ha^{-1}
iLUT	pasture	0
from 0	pine-0	0
until 10	pine-10	760.19
from 10	eucalypt-0	0
until 20	eucalypt-10	1758.58
from 20	pasture	0
until 30	pasture	4356.47

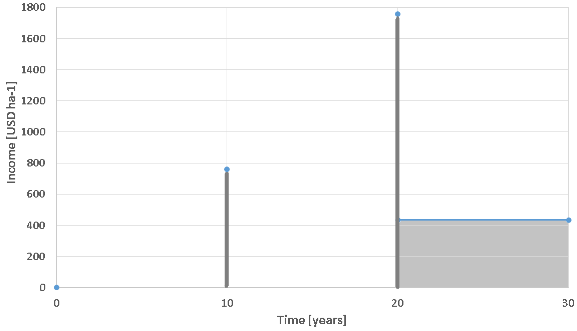

Table 6.6: Monetary income values and graph for sample land unit and land use trajectory. Total monetary income over 30 years: 6875 USD ha^{-1}.

that should be applied to every land unit in order to optimize the regionally-aggregated land performance of the full study area over a 30 year period.

Equations 6.1 to 6.4 correspond to the formulation of an IP model based on Composite Programming that is applied to the problem of finding a trajectory configuration that optimizes regional land performance.

$$\text{Min} \left[\lambda D + (1-\lambda) \sum_{k=1}^{q} w_k \frac{f_k^* - f_k(x)}{f_k^* - f_{*k}} \right] \tag{6.1}$$

subject to:

$$\sum_{j=1}^{m} x_{ij} = 1 \text{ for } i = 1, 2, \ldots, n \tag{6.2}$$

$$w_k \frac{f_k^* - f_k(x)}{f_k^* - f_{*k}} \leq D \text{ for } k = 1, 2, \ldots, q \tag{6.3}$$

$$x_{ij} \in \{0, 1\} \text{ for } i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, m \tag{6.4}$$

Equation 6.1 is the objective function of the Composite Programming model. It comprises the balancing term and the distance (achievement) term. More specifically:

- x_{ij} are the decision variables of the Composite Programming model. They indicate whether land use trajectory j should ($x_{ij} = 1$) or should not ($x_{ij} = 0$) be applied to land unit i.
- D is the maximum (weighted and normalized) deviation between any coordinate (criterion) of a given alternative and its counterpart in the ideal point;
- q is the number of considered criteria;
- w_k is the relative importance (weight) assigned to criterion k;
- f_k^* and f_{*k} are, respectively, the ideal and anti-ideal performance values corresponding to criterion k integrated over all considered land units in the study region;

- $f_k(x)$ is the performance value for criterion k, integrated over all land units, corresponding to the decision alternative (candidate solution) under consideration. $f_k(x)$ is given by a particular assignment of values to the decision variables x_{ij};
- λ determines whether emphasis is given to a balanced solution (λ closer to 1) or to a solution that is close to the ideal point (λ closer to 0).

Equations 6.2, 6.3 and 6.4 are the constraints of the model. The constraints in Equation 6.2 require that only a single land use trajectory j is assigned to land unit i. In these constraints m represents the total number of land use trajectories (82), while n corresponds to the number of land units (277). Therefore, the constraints in Equation 6.2 express that only 1 out of 82 decision variables corresponding to land unit i can be equal to 1, while all the other 81 variables are set to 0.

6.3.4 Sensitivity analysis

A sensitivity analysis was performed on the Composite Programming model through a series of tests outlined below. Each of them addressed a different aspect of this model:

1. Reference scenario;
2. Sensitivity to uncertainty in the input data;
3. Sensitivity to parameter settings;
4. Performance thresholds.

Reference scenario

In the reference scenario the input database detailed in Section 6.3.1 was used, λ was set to 0.5 to indicate that the same importance is given to a balanced solution and to the minimization of the distance to the ideal point, and equal weights (0.2) were assigned to all performance attributes. The output of the Composite Programming model for this scenario was used as a reference for assessing the results in the remaining phases.

Model sensitivity to uncertainty in the input data

Given the imperfections of the available database and considering all the assumptions used to handle the missing data, it was considered relevant to gain insights in the behaviour of the Composite Programming model under conditions of uncertainty in the input data. To this end, random "perturbations" were introduced in the performance curves (Tables 6.2 to 6.6). Such perturbations consisted in randomly selecting a value from a range of +/- 10% around the original breakpoint values in the curves and using this value as the new breakpoint. To perform a full perturbation, this step was repeated independently for every breakpoint in every land use trajectory for all land units in the database. Ten full perturbations were performed to produce the same number of different versions of the input database. The Composite Programming model was then executed for each of these versions and the outcomes were assessed in terms of their distance to the ideal point and in comparison to the results for the reference scenario.

Model sensitivity to parameter settings

The parameters of the Composite Programming model comprise λ and one weight (relative importance) for each of the performance attributes. At this stage, the model performance for six different values of λ in the range from 0 (exclusive emphasis on minimization of distance to the ideal point) to 1 (exclusive emphasis on obtaining a balanced solution) was tested, using a step of 0.2, i.e., λ was set successively to 0, 0.2, 0.4, 0.6, 0.8 and 1, while keeping all weights fixed at a value of 0.2. To assess the sensitivity of the model to the weight parameters, five separate tests were performed. In each of these tests, the weight assigned to one performance attribute was set to 0.6 while the weights for all other attributes were set to 0.1. In these tests the value for λ was fixed to 0.5.

Performance thresholds

The Composite Programming model can accommodate requirements for one or more performance attributes (not) to exceed certain predefined threshold values. Such requirements can be expressed in the form of additional constraints. Since in this model (Equations 6.1-6.4) $f_k(x)$ represents the performance value for attribute k integrated over all land units, a performance threshold constraint takes the form $f_k(x) \leq u$ for an attribute to be minimized, or $f_k(x) \geq u$ for an attribute to be maximized, where u represents a user defined threshold. In this specific case, monetary income and carbon stock were chosen to illustrate the use of such performance thresholds.

6.4 Results

6.4.1 Ideal and anti-ideal points

The first step in the Composite Programming method is to compute the ideal and anti-ideal points. The ideal point is a hypothetical decision alternative for which the coordinate value of each of its attributes corresponds to the absolute optimum that such attribute can reach, that is, each of the coordinates of the ideal point is obtained by optimizing the corresponding attribute independently from the others, thus the conflict among attributes is neglected. In this case the ideal point coordinates were computed by determining, for each land unit, the trajectories resulting in (i) minimum runoff, (ii) minimum sediment, (iii) maximum SOC, (iv) maximum BOC and (v) maximum income. Then, these optimal performance values for each attribute was summed up over all land units in order to compute the regionally integrated ideal point coordinate. The anti-ideal point was computed in a similar way, only that the anti-optimal values were chosen for each combination of land unit and attribute. The resulting ideal and anti-ideal coordinates are listed in Table 6.7.

6.4.2 Reference scenario

The settings for the reference scenario involve using the original input database, with $\lambda = 0.5$ (equal emphasis on minimization of distance to the ideal point on the one hand and balanced solution on the other hand), and assigning the same relative importance to each of the attributes (all weights set to 0.2).

Attribute	Ideal value	Anti-Ideal value	Unit
Runoff	72,611,012	772,071,110	m^3
Sediment	95,675	1,733,716	ton
SOC	1,260,563	77,853	ton
BOC	813,446	0	ton
Income	214,059,996	-6,378,626	USD

Table 6.7: Ideal and anti-ideal points for the Tabacay database. Values are aggregated at regional scale over a 30 year period.

The distribution of land use trajectories that optimizes land performance resulting from the application of the Composite Programming model for the reference scenario consisted in the assignment of 27 out of 82 trajectories over the considered land units. Table 6.8 lists the seven trajectories in this distribution that cover the largest area in the Tabacay catchment.

The first land use trajectory in Table 6.8 indicates that pasture is implemented at year 0 and kept until year 10, when it is replaced by a eucalypt forest, which is kept until year 20, when the 10 years old eucalypt forest is replaced by crops, which is in turn kept until the end of the 30 year time span. A similar interpretation can be made for the remaining land use trajectories in the first column of Table 6.8. The second and third columns of this table are the percentage of the full catchment area (excluding land units under páramo and natural vegetation) and the number of land units in which the corresponding trajectory should be implemented according to the output of the Composite Programming model. Figure 6.3 shows the spatial distribution over the Tabacay catchment of the land use trajectories listed in Table 6.8.

In Table 6.8 and Figure 6.3 a relative predominance of forest trajectories over crops and pasture is apparent. This fact is as expected since, in general, pine and eucalypt forests result in higher performance levels for most of the attributes taken into account. The extreme case is BOC, which is assumed to be always 0 for crops and pasture. Higher performance levels for pine and eucalypt are also observed in the case of runoff, sediment and SOC. On the other hand, income values are typically higher for crops and pasture, which probably explain the few occurrences of these LUTs among the trajectories that cover most of the

Land use trajectory	Area covered [%]	# land units
pasture → eucalypt-0 → crops	27.7	25
pine-0 → pine-10 → pine-20	23.8	80
eucalypt-0 → eucalypt-0 → pine-0	12.3	4
pasture → eucalypt-0 → pine-0	12.0	5
eucalypt-0 → eucalypt-0 → eucalypt-0	10.8	24
crops → crops → crops	3.1	54
pasture → pine-0 → eucalypt-0	2.9	7
Total	92.6	199

Table 6.8: Distribution of land use trajectories over the Tabacay catchment that optimizes regional land performance, obtained by applying the Composite Programming model for the reference scenario (7 out of 27 assigned trajectories are reported).

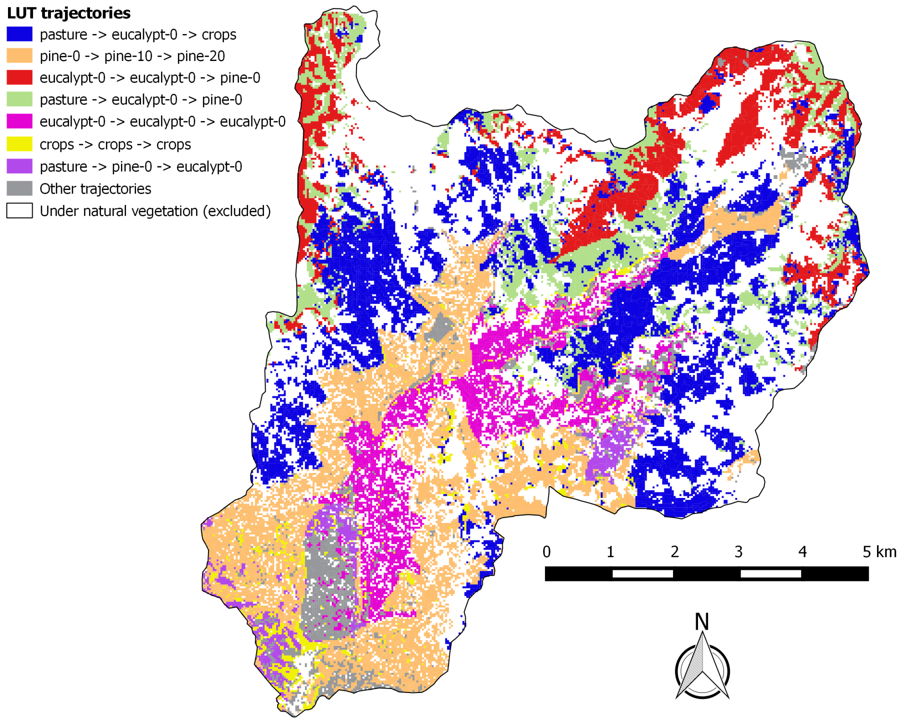

Figure 6.3: Distribution of land use trajectories over the Tabacay catchment that optimizes regional land performance, obtained by applying the Composite Programming model for the reference scenario. The category 'Other trajectories' in the legend comprises 20 different trajectories that were assigned by this model to relatively small areas.

Tabacay area.

The normalized distance to each coordinate of the ideal point can be used as an indicator of the 'quality' of a solution produced by the Composite Programming model. This indicator is a continuous value ranging between 0, when the coordinate of the solution is equal to that of the ideal point, and 1, when the coordinate of the solution is equal to that of the anti-ideal point (second term of the Composite Programming model's objective function, Equation 6.1). The normalized distances to the ideal point corresponding to the solution for the reference scenario are listed in Table 6.9.

It can be seen in Table 6.9 that the solution corresponding to runoff and income for the reference scenario is halfway between the anti-ideal and the ideal points, while the normalized distances for the other attributes are smaller, which is particularly noticeable in the case of SOC, for which the reference solution is less than 10% away from the ideal.

The existence of multiple optimal solutions depends on the specific characteristics of the problem and its corresponding formulation as a Mathematical Programming model. However the large quantity of degrees of freedom (277 land units, 82 possible land use trajectories, 5 performance attributes) in the studied case makes the existence of multiple optimal configurations of land use trajectories (for the same scenario) very unlikely. On the other hand, there may be several solutions that are close to the optimal one. To investigate the possible existence of close-to-optimal configurations we first computed the least optimal solution, i.e., the land trajectory distribution which results in a regional performance that is closest to the anti-ideal point. Next we assessed trajectory configurations of which we hypothesized that they achieve close-to-optimal ecosystem service delivery at the basin scale. The first one corresponds to selecting the land unit with the smallest area (0.09 ha) and replacing the land use trajectory assigned for optimality (i.e., pine-0 $\rightarrow$ pine-0 $\rightarrow$ pasture) by an arbitrary trajectory (agricultural land use over the full time span, crops $\rightarrow$ crops $\rightarrow$ crops). A similar approach was taken for the second trajectory configuration but here the land unit occupying the largest area (210.6 ha) in the Tabacay basin was selected and its optimal trajectory replaced by crops $\rightarrow$ crops $\rightarrow$ crops. The full performance range is assessed as the Euclidean distance between the normalized coordinates of the optimal solution for the reference scenario ([0.488, 0.228, 0.907, 0.685, 0.512]) and their counterparts for the least optimal (closest to anti-ideal point) solution ([0.671, 0.672, 0.026, 0.264, 0.328]). The coordinates for both the optimal and least-optimal solutions are normalized in the range [0, 1] with respect to the corresponding coordinates of the anti-ideal and ideal points. They must not be confused with the normalized distances to the ideal point (Table 6.9). The normalized Euclidean distance between the

Attribute	Normalized distance
Runoff	0.488
Sediment	0.228
SOC	0.093
BOC	0.315
Income	0.488

Table 6.9: Normalized distances to the ideal point corresponding to the optimal distribution of trajectories produced by the Composite Programming model for the reference scenario.

optimal and least optimal solution equals 1.104 (unitless, [-]). The distance between the optimal solution and the first disturbed configuration equals 0.0005 [-] while the second disturbed configuration is at a distance of 0.0397 [-] from the optimal solution. The latter two distances represent 0.05% resp. 3.59% of the range and indicate indeed that land use trajectory configurations exist that deliver ecosystem services to almost the same extent as the optimal one.

6.4.3 Model sensitivity to uncertainty in the input database

This section reports on the effects that controlled changes in the input database produced on the output of the Composite Programming model. In particular, this model was executed 10 times, using a different version of the input database in each test. To produce the modified input database for each run, a random perturbation was introduced in the data.

The indicator used to compare the level of agreement between the output of the Composite Programming model for two different scenarios is the coincidence index and it is expressed as the ratio between the number of land units for which both runs suggested the same land use trajectory and the total number of land units under consideration. The coincidence index is a value between 0 (no agreement for any land unit) and 1 (agreement for all land units). The coincidence index was computed to determine the agreement between the output of the Composite Programming model for the reference scenario and the results of each of the 10 runs corresponding to the stochastically modified versions of the input database. The coincidence index for these 10 tests ranges from 0.68 to 0.76 with an average value of 0.72.

In an extreme case, a land unit might be assigned a different trajectory in each of the 11 tests (one test for the reference scenario and ten for the randomly modified databases). On the other hand, it is possible that the same trajectory is assigned to a given land unit in all tests. Table 6.10 lists the number of land units that were assigned each possible number of distinct trajectories.

Table 6.10 shows that 106 out of 277 land units were assigned the same trajectory in all tests. Of the remaining land units, 55 were assigned two distinct trajectories. In general most land units were assigned a relatively low number of distinct trajectories throughout all tests, which is an indicator of the limited impact that the imposed variations of the input data have on the results of the Composite Programming model.

6.4.4 Model sensitivity to different parameter settings

The impact that different parameter settings have on the results produced by the Composite Programming model is reported in this section. Before running this model, the values for a number of parameters must be set, namely, λ and one weight for each considered attribute.

To evaluate the model sensitivity to λ, the original version of the database was used as input and the weights for all attributes were fixed to 0.2. Then the Composite Programming model was executed six times, varying λ from 0 (full priority to the minimization of the distance to the ideal point) to 1 (full priority to a balanced solution) in 0.2 steps. Table 6.11 lists the coincidence index when comparing the output of these six tests with the output for the reference scenario.

Taking apart the case in which a fully balanced solution was required, the values for the coincidence index for all other cases are relatively high, which indicates a limited sensitivity of the Composite Programming model with respect to the λ parameter. The low coincidence value in case of a fully balanced solution follows from the fact that the balance level of the reference scenario solution is rather limited, which is exemplified in Table 6.9 by the difference between the distances corresponding to runoff and income (slightly below 0.5) on the one hand, and the one for SOC (< 0.1) on the other hand.

Distinct trajectories	# land units	Percentage
1	106	38.3
2	55	19.9
3	47	17.0
4	35	12.6
5	15	5.4
6	7	2.5
7	4	1.4
8	3	1.1
9	3	1.1
10	1	0.3
11	1	0.3
Total	277	100

Table 6.10: Number of distinct trajectories assigned in 11 tests of the Composite Programming model sensitivity to variations in the input data. The corresponding land unit count is shown in the second column.

To illustrate the influence of the λ parameter on the balance of the resulting solution, Table 6.12 reports the normalized distances for the scenarios corresponding to $\lambda = 0$ and $\lambda = 1$.

To evaluate the impact of different weight settings on the output of the Composite Programming model, a set of five tests was performed. In each test the weight corresponding to one of the performance attributes was set to 0.6, while all other weights were fixed to 0.1. λ was set to 0.5 in all cases. Table 6.13 lists the coincidence index values for these tests when comparing their output to the outcome for the reference scenario.

In general the coincidence values in Table 6.13 are lower than the ones corresponding to the sensitivity tests for λ. This behaviour can be explained by the fact that, in each of the tests reported in Table 6.13, all five weight parameters were changed with respect to the reference scenario, while for the tests involving the sensitivity to λ, a single parameter was adjusted in each case. In this contexts it must be acknowledged that when a higher weight is given to an attribute with a low normalized distance for the reference solution,

λ value	Coincidence index
0.0	0.76
0.2	0.84
0.4	0.94
0.6	0.92
0.8	0.73
1.0	0.39

Table 6.11: Coincidence index values comparing the output of the Composite Programming model for the reference scenario and for scenarios using different values for the λ parameter (all weights were fixed to 0.2).

Attribute	Normalized distance	
	$\lambda = 0$	$\lambda = 1$
Runoff	0.318	0.445
Sediment	0.205	0.250
SOC	0.093	0.442
BOC	0.259	0.445
Income	0.685	0.445

Table 6.12: Normalized distances of the solution produced by the Composite Programming model for $\lambda = 0$ and $\lambda = 1$ (all weights were fixed to 0.2).

the correspondence with the reference configuration of trajectories decreases. Consider for example the case in which more importance is given to runoff (reference normalized distance = 0.488). In that case, less than 40% of the land units are assigned the same trajectory with respect to the reference scenario. On the other hand, if emphasis is given to SOC (reference normalized distance = 0.093), almost 70% of the land units are assigned the same trajectory as for the reference. As further illustration, Table 6.14 lists the normalized distances for the solution produced by the Composite Programming model when the weight for runoff is set to 0.6 and all other weights are fixed to 0.1.

6.4.5 Performance thresholds

The possibility of accommodating performance thresholds in the Composite Programming model was tested using the attributes monetary income and carbon stock in the soil and in the biomass. The first performance threshold constraint was defined to require income to be at least as high as 10% of the minimum amount of money that is required to fulfil the basic needs of all

Weight setting	Coincidence index
$weight_{runoff} = 0.6$, all other weights = 0.1	0.39
$weight_{sediment} = 0.6$, all other weights = 0.1	0.54
$weight_{SOC} = 0.6$, all other weights = 0.1	0.69
$weight_{BOC} = 0.6$, all other weights = 0.1	0.51
$weight_{income} = 0.6$, all other weights = 0.1	0.35

Table 6.13: Coincidence index values comparing the output of the Composite Programming model for the reference scenario and for scenarios using different weight settings (λ was set to 0.5).

Attribute	Normalized distance
Runoff	0.042
Sediment	0.144
SOC	0.485
BOC	0.394
Income	0.901

Table 6.14: Normalized distances of the solution produced by the Composite Programming model for $weight_{runoff} = 0.6$ and all other weights fixed to 0.1 (λ was set to 0.5).

inhabitants in the Tabacay catchment. To compute the income threshold value several information elements were considered. The total population of Tabacay for every year in the period 2015-2045 was estimated using census information available for 2010 and the population projections for the period 2011-2020[1]. Furthermore, information about the minimum income required by a family in the province of Cañar to fulfil its basic needs was retrieved as monthly values for the current and past years. Using this information as a reference, the yearly income required by a family in Tabacay was estimated for each year the period 2015-2045. The total income required to cover the basic expenses of the full population of Tabacay for the period 2015-2045 was then computed using Equation 6.5.

$$u_{income} = \sum_{y=2015}^{2045} f_y i_y \tag{6.5}$$

where u_{income} is the value for the income threshold, f_y is the number of families living in Tabacay at year y (an average family size of four was assumed), and i_y is the minimum income required by a family in Tabacay at year y.

From Equation 6.5 the minimum amount of money required to fulfil the basic needs of the population of Tabacay is above 1.58 billion USD, whereas the income coordinate of the ideal point (maximum income possible) is around 214 million USD. This means that, when considering only the income from crops, pasture, pine plantation and eucalypt plantation in Tabacay, at most 13.6% of the total amount of money required to cover the basic needs of the population can be generated. The performance constraint for income considered a threshold of 158 million USD, i.e., 10% of the value resulting from Equation 6.5.

From Table 6.9, it can be seen that the solution generated by the Composite Programming model for the reference scenario results in a SOC stock deviating 9.3% and a BOC stock deviating 31.5% from their corresponding ideal values. Since the solution performance regarding SOC was considered relatively high, a performance threshold constraint was included in the Composite Programming model to require that such level is maintained (SOC $\geq$ 1,150,571 ton). On the other hand, a performance threshold was defined to require that the deviation from the ideal concerning BOC improves from 31.5% until at least 25% (BOC $\geq$ 610,000 ton).

The resulting (partial) distribution of trajectories over Tabacay after applying the Composite Programming model to the reference scenario and considering

[1] `http://www.ecuadorencifras.gob.ec/proyecciones-poblacionales`, in Spanish, last accessed on July 17th, 2015

the performance threshold constraints described above is summarized in Table 6.15. The corresponding spatial distribution is displayed in Figure 6.4.

In Table 6.15 and Figure 6.4 there is a clear predominance of forest land use trajectories, especially trajectories involving eucalypt plantation. This fact is as expected, considering the thresholds set for carbon stocks and that non-forest LUTs are assumed to always have a BOC of 0. The frequent appearance of eucalypt follows leads us to conclude that this species presents a good performance regarding carbon sequestration and income. On the other hand, the appearance of the trajectory involving an agricultural land use during the full 30 year time span may be explained by the requirement to fulfil the income threshold. Table 6.16 reports the normalized distances to the ideal point corresponding to the scenario with income and carbon stock performance thresholds.

Land use trajectory	Area covered [%]	# land units
pine-0 → pine-10 → pine-20	25.1	73
eucalypt-0 → eucalypt-0 → crops	20.0	12
eucalypt-0 → eucalypt-0 → pine-0	18.0	7
crops → crops → crops	16.6	98
eucalypt-0 → eucalypt-0 → eucalypt-0	13.3	36
Total	93.0	226

Table 6.15: Distribution of land use trajectories over the Tabacay catchment that optimizes land performance, obtained by applying the Composite Programming model for the reference scenario and considering income and carbon stock thresholds (5 out of 25 assigned trajectories are reported).

Attribute	Normalized distance
Runoff	0.805
Sediment	0.343
SOC	0.093
BOC	0.250
Income	0.255

Table 6.16: Normalized distances of the solution produced by the Composite Programming model when considering income and carbon stock thresholds.

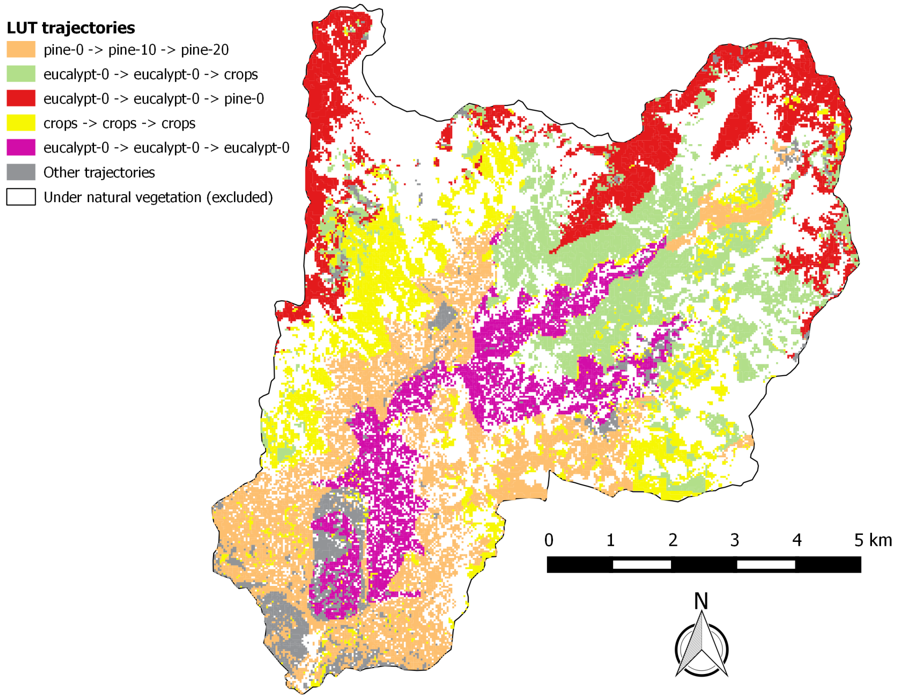

Figure 6.4: Distribution of land use trajectories over the Tabacay catchment that optimizes land performance, obtained by applying the Composite Programming model for the reference scenario considering income and carbon stock thresholds. The category 'Other trajectories' in the legend comprises 20 different trajectories that were assigned by this model to relatively small areas.

6.5 Discussion

The general aim of the Composite Programming model studied and applied in this chapter is to determine how a set of LUTs should be distributed over space and time in order to optimize the multi-dimensional regional performance of land. In this particular case, this model was applied to a database that consists of a set of land units representing the catchment of the Tabacay river, located in Ecuador. This database contains values corresponding to five performance attributes for each possible combination of land units and LUTs: runoff production, sediment production, stock of SOC, stock of BOC, and monetary income. These performance values represent the state of affairs in the catchment at a reference point in time (year 0), and 10, 20 and 30 years after.

Four different LUTs (crops, pasture, pine plantation and eucalypt plantation) and 277 land units are considered. Any of these LUTs can be applied to a land unit either at year 0, 10, or 20. A land use sequence, or trajectory, is defined by the combination and order in which these LUTs are established. the Composite Programming model was applied in this study to determine the land use trajectory, over a 30 years time span, that should be implemented in each land unit in order to optimize the regional land performance expressed in terms of the five attributes listed above.

In Chapter 5 the Composite Programming model was applied to solve a problem that involved determining the distribution of a number of LUTs over the Tabacay catchment in order to optimize the multi-dimensional land performance at a regional scale. In that chapter it is assumed that the LUT assigned to any land unit is kept during the full period of 30 years. In this sense, the land use configurations resulting from the application of the Composite Programming model in Chapter 5 are static. In the present chapter, the application of that model was taken one step further, by incorporating the temporal dimension. The restriction of allowing a land use change to occur only at the beginning of the considered period is no longer present. Instead, land use changes may (but must not) occur at 10 year intervals within the 30 year period. From this perspective and given the definition of trajectory, which implicitly suggests the possibility of land use changes after preset intervals, the application of the Composite Programming model in the present study can be considered a generalization of the work reported in Chapter 5. Although time was discretized into relatively coarse periods, this work can be considered as a first step towards more flexible models that allow to tackle a broader category of problems.

An apparent predominance of forest LUTs was observed among the land use trajectories suggested by the Composite Programming model for the Tabacay catchment. This fact is a result of the superior performance of forest over crops and pasture, which is particularly evident for the case of BOC, but also notorious for runoff and sediment production and for SOC. On the other hand, land performance corresponding to agriculture and pasture exceeds forest in terms of income, which is the reason why non-forest LUTs are sporadically present in the solution provided by the Composite Programming model.

It is obvious, however, that the fact that the values for monetary income were not corrected for devaluation over time by applying a discount rate has an impact on the absolute results, i.e. the location in feature space of the ideal point and the distance to the ideal point of the optimal (and other) land use trajectory configurations. For example, a constant yearly discount rate of 5% would make the income coordinate of the ideal point shift from an assumed 1000 USD to 950 USD after one year, 598.70 USD after 10 years, 358.50 USD after 20 years and 214.60 USD after 30 years. Moreover, the discounting makes a single LUT selected as the second component of a land use trajectory to have a lesser absolute contribution to the overall land performance than when it is selected as the first component and a higher contribution with respect to when it is selected as the third component. This behaviour is not applicable though to the four other ESS considered in this study. However, since a same discount rate would be applied to the income values generated from all LUTs that can be part of a trajectory, the nature and sequence of LUTs in the selected trajectories will mostly not be affected.

In particular, given a specific trajectory configuration resulting from the application of the Composite Programming model to a given scenario in which the discount rate was not taken into account, the only trajectories that would be significantly affected, when discount rates are applied, are the ones that contain a high performing LUT in terms of income (e.g., crops or pasture) during the second or third time period (i.e., from year 11 to 20 or 21 to 30). Such trajectories will be presumably replaced by others in which high income LUTs appear in the first or second intervals of the time span (i.e., 0-10 or 11-20). This is based on the fact that such changes do not affect in a considerable way the performance attributes other than income. For instance, considering the sample land unit, for which its performance is illustrated in Tables 6.2 to 6.6, if pasture (which performs better in terms of income than pine and eucalypt) was established in the first time interval (0-10) instead of the final interval (21-30), such trajectory change would not affect the total runoff produced during the full time span. The same can be said for the case of sediment and BOC. On the other hand, for trajectories involving cutting and replanting forest of different tree species in consecutive periods, runoff and sediment values will effectively change when the trajectory is modified in such a way that the periods under forests are not consecutive any more or the tree species used for the consecutive periods are swapped.

Applying the reasoning described above to the solution for the reference scenario described in Table 6.8, one can argue that the only trajectory affected when applying discount rates would be the first one (pasture → eucalypt-0 → crops). In that hypothetical case, this trajectory might be replaced by one in which crops appears earlier, possibly in the first time interval. The last trajectory in Table 6.8 (pasture → pine-0 → eucalypt-0) might also be replaced by one in which pine and eucalypt are swapped, given the higher performance in terms of income that is normally found in eucalypt forest over pine plantations. The other trajectories in Table 6.8 would presumably remain the same, although the specific land units to which it is applied and, thus, the area covered by each of the trajectories might change. In summary, some of the selected trajectories will change when considering discount rates in income, although such changes will not have a major effect on the overall trajectory configuration suggested by the Composite Programming model for the whole study region.

Ten different tests of the Composite Programming model were performed, using at each time a randomly modified version of the original database, with the aim of analyzing the behaviour of this model under simulated conditions of uncertainty in the input data. In all these tests, around 70% of the land units were assigned the same trajectory as the case in which the original database was used. This fact indicates a relatively low sensitivity of the output of the Composite Programming model with respect to imposed variations in the input database, considering that all performance values were randomly modified within a restricted range (+/- 10% of the original performance value). It is indeed expected that the similarity in the results decreases when the range of variation is enlarged. Another indicator of this sort of stability is that around 38% of the land units were assigned the same trajectory in all tests, with an additional 20% being assigned only two distinct trajectories. In general, most land units were assigned a limited number of distinct trajectories when using different versions of the original database.

In addition to the sensitivity of the Composite Programming model to simulated uncertainty in the input data, the impact of different parameter settings on its results was also assessed. The functioning of this model depends on several parameters. The first parameter is λ, which indicates whether the priority is to find the solutions that are the closest to the absolute optimal (λ closer to 0) or a solution that achieves a balanced level for all performance attributes under consideration (λ closer to 1). In addition to λ, one weight parameter is provided for each performance attribute to denote the relative importance that such attribute has for the decision maker. To evaluate the sensitivity of the Composite Programming model to different parameter settings, several tests were executed varying λ and the set of weights independently within their full allowed range ([0, 1]).

The changes on the results when varying λ between 0 and 0.8 were confined to a range of 25% with respect to the case when λ was set to 0.5, which indicates that the sensitivity of the Composite Programming model to λ is relatively restricted, even for drastic variations. When full balance of the solution was required ($\lambda = 1$) the difference with regard to the reference increased to 60%, which means that the fully balanced solution is rather far from the reference in the solution space. This fact is clearly revealed when contrasting Tables 6.9 and 6.12. With $\lambda = 1$ (balanced solution is an absolute requirement), the overall deviation from the ideal point is largest, followed by $\lambda = 0.5$ (reference scenario) and $\lambda = 0$ (exclusive emphasis in minimization of distance to ideal point, balanced solution is not required). This is in line with our expectations that a strictly balanced solution is not preferred since balancing interferes with the phenomenon of trade-off among the considered attributes. The expected stock of soil carbon (SOC) is particularly sensitive to λ as its deviation from the ideal point is ranging between 9.3% (for $\lambda = 0$) and 44.2% (for $\lambda = 1$). Moreover it is remarkable that SOC is the only performance attribute for which the deviation for $\lambda = 0.5$ is equal to the one obtained for $\lambda = 0$. It is interesting to observe the level of conflict existing between runoff and income and the way in which such conflict is traded off by the Composite Programming model. Concretely, for all tests performed, when a solution resulted in a runoff level close to the ideal point (e.g., Table 6.14), the corresponding income level was far from its ideal value, and vice versa (Table 6.12, when $\lambda = 0$). When one of the attribute levels is halfway between the anti-ideal and the ideal points, the other one shows a similar behaviour (Table 6.9). In general, the explanation for the described behaviour of all performance attributes is not simply ecological but has more to do with the large number of possible combinations of land units with LUTs and their sequence.

Differences in the output of the Composite Programming model were more notorious when varying the weight parameters, presumably due to the fact that not just one but a set of several parameters (one per attribute) were modified in this case. As reported in Table 6.14, setting a weight of 0.6 for runoff and of 0.1 for each of the four other criteria results in relatively large deviation and makes SOC and especially income deviate more from the ideal point than in the three other scenarios ($\lambda = 0$, $\lambda = 0.5$ and $\lambda = 1$) in which the weights for all criteria was equal (0.2). Income from agricultural activities is typically higher compared to forestry while the opposite is true for other four attributes. Hence it is not surprising that income shows a large deviation from the ideal point in case one of the four other attributes has a dominant weight. This is especially true in the case of assigning a larger weight to runoff, given the degree of conflict between this attribute and income.

It is important to note that the formulation of the Composite Programming model, or any other model of this kind, as an IP model relies on the assumption that each land unit was defined, with as much certainty as possible, as a separate patch that should be managed according to a single land use trajectory. To guarantee that this option is the most convenient approach, the land use planner must make sure that all the relevant information about land characteristics was taken into account when defining the land units. Determining the different criteria that will be used to define land units can be a challenging endeavour, since several uncertainties and subjectivities can be involved in the process.

A more flexible approach would be to formulate the Composite Programming model as a LP model (Winston and Goldberg (2004)). This would imply adapting this model's constraints in such a way that its decision variables can take fractional values in the range [0, 1]. The solution of this variant would express that the corresponding fraction of the land unit at stake is recommended to be managed according to a certain trajectory. This LP variant of the Composite Programming model would then allow for several trajectories to be assigned to different fractions of a single land unit. This would probably result in a more realistic output, in which land units are not necessarily devoted to a single trajectory. Although this LP variant would be able to indicate the fraction of a land unit that should be managed according to a specific trajectory, it would not give any information about the particular location within the land unit in which the trajectory should be applied.

Although in this particular study the requirements regarding execution time when solving the Composite Programming model were not highly demanding, there could be situations, depending on the size and other particularities of the input data, in which execution time becomes a limiting factor. In such cases, the LP variant of this model would clearly stand out as the preferred option, given that LP models in general are solved in radically shorter times when compared to their IP counterparts. Another advantage of LP models is that their solutions are an upper bound of their IP counterparts. This means that the quality of the solutions produced by a LP model is guaranteed to be *at least as high* as solutions generated by IP. As a matter of fact, in most cases the solution produced by a LP model performs better than its IP counterpart. On the disadvantages side, when the Composite Programming model is formulated as a LP model, there is the risk of obtaining solutions that are excessively fragmented, that is, solutions in which many trajectories are assigned to every land unit. Such solutions can be difficult to apply or even infeasible in reality. However, in this particular study, the solutions produced by the IP and LP versions of the Composite Programming model did not differ at a large extent. The risk of over-fragmentation is not present in the IP formulation when the land units have been defined in a sensible way.

Finally, it is important to mention that, in the applications of the Composite Programming model reported here and in Estrella et al. (2014a), the environmental performance of a land unit is assumed not to be influenced by the state of other land units. However for several land use planning problems the consideration of spatial interaction between a land unit and neighbouring or even distant land units is pertinent (Vanegas et al. (2012); Estrella et al. (2014c)). A performance attribute that is influenced by the notion of spatial interaction is said to be an instance of an off-site attribute. An initial attempt of incorporating an off-site attribute in the formulation of a Mathematical Programming model is reported in Vanegas et al. (2012) for the specific case of sediment delivery to the river in a catchment. A more restricted example of a tool that implements the notion of spatial interaction is Holzkämper and Seppelt (2007), in which each cell of a raster representing the study region is assigned several attributes. These attributes can be expressed as a function either on local characteristics or on characteristics of its neighbouring cells. Holzkämper and Seppelt (2007) do not consider the influence that distant cells may have on the state of a given cell. In Schädler et al. (2012) several assessment methods to analyse and visualize site specific spatial information are described. These methods are aimed at deriving and recommending mainly urban land use layouts to redevelop degraded areas. Preference for the assignment of a certain LUT to a spatial unit is given by, among other factors, the LUTs present in its neighbouring areas. In a closely related study (Schädler et al. (2013)) a set of sustainability indicators is derived with the aim of matching planning units to LUTs. Some of these indicators are given by the LUT present in surrounding areas, e.g., whether residential areas or green spaces are present in the surroundings, presence of commercial areas within walking distance, among others.

6.6 Conclusions

This chapter presents a transparent, flexible and robust Composite Programming model for determining an optimal land use distribution in discretized space and time, taking multiple criteria into account. The performance of this model was evaluated using quantifiable indicators, like the normalized deviation from the ideal point and the coincidence index. Levels of performance attributes are used as criteria so that the Composite Programming model promotes land use plans that can serve as a canvas for further specific planning. The formulation of this model makes it applicable to any problem in which the aim is to determine how to use land over time with a view to optimize the integrated performance of a given region, when such performance is expressed as a number of continuous attributes.

The outcome of the optimization exercises for the Tabacay river basin in the southern Andes of Ecuador, as discretized in 277 land units and taking four LUTs (arable land, pasture and plantation of eucalypt and pine trees) and five on-site ecosystems services into account (runoff production, sediment production, SOC sequestration, organic carbon sequestration in biomass and monetary income, generated from the LUTs and land use trajectories on the applicable land units), leads to two major conclusions: (i) the Composite Programming model allows to obtain stable solutions even when the validity of the input data is not absolutely guaranteed and (ii) the phenomenon of trade off between the provisioning service of the land (income) with the regulation and maintenance services (runoff, sediment, SOC) is crucial.

The Composite Programming model succeeds in accounting for the complex multi-dimensional phenomenon of trade-off between ecosystem services when determining the optimal spatio-temporal distributions of LUTs. Despite this complexity, the expectation that the weights attributed to the provisioning or to the regulation and maintenance services are the main determinants for having the land use distributions dominated by either agriculture or forest is confirmed.

Chapter 7

Conclusions and perspectives

7.1 Conclusions

In this dissertation the problem of land use planning in general and afforestation planning in particular has been tackled from several distinct perspectives. An afforestation planning problem, as defined in this study, involves several dichotomies: (i) single or multiple criteria; (i) on-site or off-site criteria; and (iii) optimization 'per land unit' or at a regional scale. The approach chosen to solve a concrete afforestation planning problem will depend on which facet of each of these dichotomies is addressed. This dissertation is focused on covering a few problem instances from among all the combinations that can be possibly stated according to this framework (see Figure 7.1).

In a first part, a site location problem involving the optimization of a single off-site criterion at a regional scale is addressed. The second part is also targeted to solve a site location problem, although in this case the aim is to optimize multiple on-site criteria on a 'per land unit' basis. In the last part, the focus is shifted from site location problems to land use management while considering the study region as an integral entity. At this point, several MCDM methods are applied to determine the way in which a number of predefined LUTs should be distributed over a region in order to obtain an optimal land performance at a regional scale after a given period of time. Land performance is expressed as the trade-off of multiple on-site criteria, and both a static (no land use changes are considered during the time period) and a dynamic (land use changes may or may not occur) management scenarios are considered.

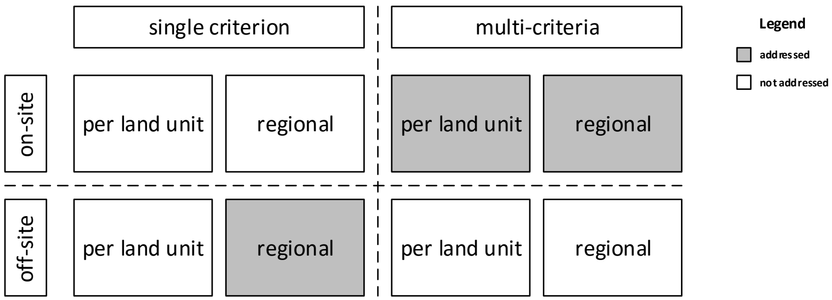

Figure 7.1: Aspects of afforestation planning that have and have not been addressed in this dissertation.

The sections below give more details about how each of these problem instances was addressed.

7.1.1 Locating afforestation sites to optimize a single off-site criterion

Site location is a fundamental issue in land use planning in general and afforestation planning in particular. The question of which spatial units, either raster cells or land units, are best suited to be covered by a given LUT has been addressed in the past using techniques taken from the discipline of land evaluation. Vanegas et al. (2012) proposed a site location technique called Cellular Automata based method for Minimizing Flow (CAMF) to select, from a rasterized representation of a river catchment, a predefined number of cells that should be afforested in order to minimize sediment yield. In this context, the term sediment yield is an off-site attribute that is defined as the amount of sediment that is delivered to the outlet of a river catchment.

Several performance aspects of CAMF were assessed, in particular its behaviour when dealing with large datasets and the impact of the number of cells to be selected, i.e., the solution size, on its execution time. During these assessments, two main issues were identified on the internal working of this method. Firstly, CAMF is an iterative method that ranks all cells in its input dataset according to their potential for sediment yield reduction when covered by forest. The computation of such potential for each cell is performed through the execution of a sediment flow simulation algorithm. This algorithm, though simple, is nevertheless expensive in terms of computational resources usage. This fact is exacerbated in cases in which the number of candidate cells to be afforested is large, i.e., when the input dataset covers a large area and/or the spatial resolution of the dataset is high. The second weakness of CAMF is that it selects a limited number of optimal cells in each iteration. This characteristic makes the execution of many iterations necessary in order to fulfil the required solution size. The combination of these two issues, that is, that CAMF selects only a few cells from a costly-built ranking at each iteration can result in the risk that the number of cells in the input dataset and the number of cells required to be afforested become limiting factors for the applicability of CAMF. More specifically, the application of CAMF can prove infeasible for problems that involve choosing a large number of cells to be afforested while using high resolution input datasets that cover an extensive study area. This restriction can be even more critical in contexts where CAMF needs to be applied repeatedly in a systematic way, as it could be the case when performing scenario analysis, or when using CAMF as a component in an integrate method that requires to execute it in an iterative fashion.

These challenges led to the hypothesis that improvements can be introduced in CAMF in such a way that its efficiency is significantly increased while keeping a comparable quality in terms of results. With the aim of alleviating the limitations of CAMF stated above, a variant of CAMF, called local CAMF was proposed and applied. This variant uses only local cell information pertaining to sediment reduction due to afforestation and slope for ranking cells. This adaptation made possible to select all cells that should be afforested in a single step, avoiding the requirement of performing many iterations. Local CAMF proved to produce very similar results with respect to CAMF, requiring only an almost negligible, constant execution time.

A fundamental notion applied in CAMF when simulating sediment flow within a catchment is the concept of spatial interaction. This concept refers to the fact that changes in the state of a cell may have an impact on the state of neighbouring or even distant cells. The fact that in local CAMF spatial interaction is completely neglected, and that it still produces very similar results with respect to the original version of CAMF, can be indicators that the notion of spatial interaction is not as relevant as it was initially assumed for this kind of optimization methods. However this behaviour can also be a consequence of the ratio between the solution sizes and dataset sizes considered during the testing phase. In particular, when the number of cells to be afforested is small in relation to the number of cells in the input dataset, spatial interaction is less likely to play a role, since selected cells can be located in very disperse regions of the catchment and, hence, they do not influence each other. When the ratio solution size / database size increases, the role of spatial interaction is expected to become more apparent, and the differences in the solutions provided by the original version of CAMF and its local variant are expected to be more notorious.

In order to compute the potential of sediment reduction for each cell when afforested, original CAMF incorporates a simple sediment flow simulation component. This component is based on a SFD model. A SFD model is based on the assumption that flow leaving a cell is delivered to only one of its neighbours. A potential improvement to CAMF, already suggested in Vanegas et al. (2012), is to replace SFD by a more sophisticate MFD model. After a literature exploration, the FD8 MFD algorithm was chosen as the most convenient model to be implemented in this new variant of CAMF.

In addition to adapting the sediment flow simulation component of CAMF, a calibration procedure for all its parameters was carried out. This procedure consisted in systematically varying the parameter values within their validity ranges using fixed steps until the outcome of CAMF, in terms of sediment yield, was considered to be sufficiently close to reference results. These reference results were obtained using the well established sediment production and transport simulation model WaTEM/SEDEM (Van Oost et al. (2000); Van Rompaey et al. (2001)). The new an calibrated CAMF variant resulting from the integration of the MFD model is called CAMF-MFD. It was tested using a number of different afforestation scenarios, during which CAMF-MFD proved to be capable of producing similar results with respect to the reference, more process-based sediment production and transportation model.

7.1.2 Locating sites for afforestation to optimize multiple on-site criteria

A multitude of methods have been proposed in the past to address decision problems in which optimization is required in the light of several conflicting criteria. Some of these techniques, called MCDM methods, can be targeted to solve problems in which the set of alternatives is finite, while some of them are capable of dealing with both finite or infinite alternative sets. From the MCDM methods that deal with finite alternative sets, some of them are better suited to tackle problems in which the cardinality of this set is rather limited. That is the case for pairwise comparison and outranking methods. Other methods, while still restricted only to finite alternative sets, allow for the consideration of larger quantities of alternatives (e.g., IIPT). When the cardinality of the set of alternatives is sufficiently small, and depending on the MCDM method applied, a ranking of the alternatives can be produced according to the criteria under consideration, in descending order of suitability. This allows to select the most suitable alternatives directly from the top of the ranking. Such rankings also provide an immediate overview of the relative performance of alternatives with respect to each other.

This part of the dissertation was meant as an introduction to the field of MCDM methods. After a literature exploration, a sample comprising six frequently applied methods was chosen to demonstrate the applicability of such methods to problems requiring the location of optimal sites for afforestation. The concrete problem that was addressed was stated as to locate the sites within a river catchment that should be afforested in order to obtain an optimal land performance after a given period of time. Land performance in this case was expressed as a number of conflicting on-site criteria, both in the environmental and in the socio-economic fields. The output of each MCDM method consisted in a ranking of the 20 land units in which the catchment was initially stratified. A relatively high degree of consistency was observed when the outcomes of all MCDM methods were contrasted. This indicates that, regarding the considered criteria, the relative performance of land units remains more or less consistent no matter the method that is used to rank them. In other words, the fact that some land units outperform others in this case seems not to depend on specificities of the nature of a given MCDM method. Instead, such outperforming relationships among land units seem to be determined by the inherent characteristics of them. Additionally, the observed consistency in the behaviour of the studied MCDM methods is a sort of positive cross-validation indicator. This degree of consistency allows the decision maker to concentrate on more practical aspects when it comes to choosing a particular MCDM method to solve a given problem. Specifically, aspects like ease of use, transparency, and number of parameters to be tuned can gain a higher relevance, given the fact that the output in terms of decision support are not extremely dependant on the specific MCDM method applied. On the other hand, an absolute consistency on the outcomes of all MCDM methods cannot be realistically expected, given the very diverse nature of MCDM methods in terms of design, internal functioning and parameter settings, which can be configured in many different ways.

All MCDM methods that were tested in this part proved applicable to the sample afforestation planning problem stated. This fact put forward several methods that were considered as promising options to be applied in subsequent chapters.

7.1.3 Land use and land use trajectory configurations to optimize multiple on-site criteria

In this dissertation the term land use configuration is used to refer to the distribution of a number of predefined LUTs over a region stratified in a set of land units with a view to achieve the desired objectives. Using this definition, one can state that the ultimate outcome of land evaluation and land use planning is a concrete land use configuration. More concretely, a land use configuration consists in a specific assignment of one or more LUTs to every land unit. When land performance optimization is required, there are two different approaches to perform such assignment. The first approach focuses on the optimization of the land performance pertaining exclusively to the land unit that is under consideration. That is, this approach considers land units as independent entities that do not interact with each other, i.e., optimization is performed separately 'per land unit'. The second approach emphasizes the optimization of land performance at regional scale. In such a case, the performance levels of the individual land units are integrated over the full study region, which results in an aggregate value representing the performance of the whole region for each considered attribute. Although the difference between both approaches resides in the scale at which land performance is assessed, the outcomes of these two approaches are still comparable, since they consist in land use configurations as defined above.

In this part two methods selected from the exploration and evaluation procedure described in Section 7.1.2, namely IIPT and Compromise Programming, plus one additional MCDM method called Composite Programming, were applied to solve the problem of finding land use configurations that optimize regional land performance. All these three methods are based on the notion of the ideal point.

In a multi-criteria decision problem, every alternative can be seen as a vector in a multi-dimensional space, with each dimension corresponding to a separate criterion. The ideal point is defined as a hypothetical decision alternative for which each of its coordinates is the absolute optimum level that the corresponding criterion could reach. Given the conflict existing among criteria, the ideal point is a non-feasible alternative.

IIPT is an example of a method only suited for 'per land unit' optimization, which means that it processes land units one by one, with the process being restarted from scratch every time a land unit is assigned a LUT and a new land unit is to be processed. IIPT looks for the most suitable LUT in an iterative fashion. At each iteration step it defines a threshold as a relaxation with respect to the ideal point. The term relaxation is used here to refer to any variation introduced on the coordinates of a given threshold in such a way that the likelihood of alternatives fulfilling the new threshold is increased. After computing the threshold, the set of alternatives is checked to determine whether there exist any alternatives that satisfy it. If no alternatives are found, the threshold is further relaxed and the set of alternatives is queried again. This procedure is repeated until one or more alternatives are found to satisfy a given threshold.

IIPT was applied to every land unit to determine the LUT that should be assigned to such unit in order to optimize its multi-dimensional land performance. After all land units were processed in this way, the result was a land use configuration for the full study region. A drawback that was identified during the tests of IIPT is its incapability to distinguish among decision alternatives that present similar performance levels. Since IIPT defines at each iteration a threshold used to find high performance alternatives, multiple alternatives will be selected whenever more than one of them satisfy a given threshold. The output of IIPT in such cases is less informative.

To shift the focus from 'per land unit' to regional optimization, the Compromise Programming and the Composite Programming MCDM methods were applied. Compromise Programming is an MCDM method that ranks and/or selects decision alternatives based on their distance to the ideal point in the multi-dimensional space defined by the considered criteria. In this case the coordinates of the ideal point are calculated by summing up the optimal coordinates of all considered land units. The election of the distance measure to be used is left to the decision analyst. For instance, Euclidean distance may be used, although that would result in a non-linear formulation, which is more complex to solve than a linear problem. Composite Programming allows the inclusion of other metrics in addition to distance to the ideal point. In particular, the Composite Programming formulation proposed in this dissertation integrates elements that favour balanced solutions. In this context, the term balanced solution refers to the situation in which each coordinate of the selected alternative is uniformly close to the corresponding ideal point coordinate. Composite Programming proved to be a flexible method that allows the decision maker to fine tune the emphasis that is assigned to closeness to the ideal point and to solution balance, a feature not present in IIPT nor in Compromise Programming, which do not make explicit provisions for solution balance.

It was observed during the tests that the solutions produced by the Compromise Programming model showed some degree of balance. This indicates that, although explicit measures favouring balanced solutions are not included in Compromise Programming, the minimization of the multi-dimensional distance to the ideal point implicitly favours solutions with some degree of balance.

The application of the principles defined by both Compromise and Composite Programming result in a Mathematical (either Linear or Integer) Programming model. The formulation of Mathematical Programming models allows for a greater degree of flexibility with regard to, e.g., algorithmic methods like IIPT or CAMF. In general, the mathematical language used to formulate a decision problem in terms of an objective function and a set of constraints is very powerful, allowing to concisely express complex ideas, interactions and requirements. Adaptations to a mathematical programming model are typically restricted to the inclusion of new elements in the objective function and/or the integration of new constraints. For example, both the Compromise and the Composite Programming formulations presented in this dissertation are ready to deal with problems involving a larger or smaller number of criteria. Such change in the decision problem would require adaptations only in the input database, while the Mathematical Programming formulation would remain unaltered. As an additional example, in case that minimum thresholds need to be imposed for the area covered by a given LUT, such requirement can be easily incorporated to a Mathematical Programming model by means of a new constraint. Adaptations in algorithmic methods like IIPT are potentially harder to implement, given its more limited modularity when compared to Mathematical Programming formulations.

Although temporal considerations were implicitly addressed in previous chapters, by considering the expected land performance cumulated during or occurring after a certain time period of having established a given LUT in the study region, a more explicit inclusion of the temporal dimension is often necessary. As a first step towards a real spatio-temporal decision support method, the notion of land use trajectory was introduced that, in this context, refers to a limited number of LUTs that are applied in sequence to a land unit at fixed intervals within a time span. The start of the considered time span is set to a reference, non-absolute point in time (called year 0). To avoid excessive complexity, the concept of land use trajectory is based on the assumption that land use changes can only occur at fixed points in time, and only a limited number of LUTs are considered. The land performance of a specific trajectory is given by the multi-dimensional performance corresponding to the full time span. The assignment of a particular land use trajectory to a certain spatial unit amounts to determining the way in which land will be managed on that unit over time for the full considered period. By extension, the term land use trajectory configuration refers to a particular assignment of a trajectory to every land unit in the study region. Regional land performance for a given trajectory configuration is computed by integrating the performance values corresponding to the assigned trajectories over all land units.

Composite Programming proved to be a useful method to solve problems that require finding optimal land use trajectory configurations. Concretely, this method was applied to determine the trajectory configuration that would lead to an optimal, multi-dimensional land performance at regional scale. The output of Composite Programming for this problem can be seen as a base planning map that indicates how land use should evolve over the full region during the predefined time span.

In addition to applying the Composite Programming model to a base scenario, its sensitivity to uncertainty in the input data was assessed. Uncertainty was simulated by introducing random perturbations, within a predefined range, to the performance values in the input data. The impact of such perturbations on the results of the Composite Programming model was found to be limited, which may be an indicator of certain degree of stability on its outcome with respect to controlled variations on the input data. A relatively high level of stability was also observed when testing the Composite Programming model with different parameter settings.

The possibilities of the Composite Programming model regarding adaptability were demonstrated by including performance thresholds. A performance threshold is an aspiration expressed by the decision maker for a criterion to exceed or not to exceed a certain value, depending on whether the criterion is to be maximized or minimized, respectively. Such performance thresholds are normally given by specific conditions in the domain where an MCDM method like Composite Programming is applied. It was shown that such performance thresholds can be accommodated into the Composite Programming model by adding the appropriate set of constraints. The same approach can be applied when similar requirements like, for instance, that a specific LUT should cover a minimum area within the study region, are expressed by the decision maker.

The application of the Composite Programming model to these two related problems (determining either land use or trajectory configurations) demonstrates that when the available data can be structured in the required format, the versatility of this model allows to obtain stable solutions for different problems even when the validity of the input data is not absolutely guaranteed and when there is no complete certainty about the values used for its parameters.

7.2 Perspectives for research and development

7.2.1 Site location for afforestation to optimize a single off-site criterion

CAMF was proposed as a method for locating sites that should be afforested in order to minimize the sediment yield of a river catchment. To this purpose, CAMF incorporates a sediment flow simulation module based on the notion of spatial interaction. Spatial interaction dictates that changes in the state of a spatial unit can have an influence on the state of neighbouring or even distant units. In Chapter 2, when analysing options to shorten the execution time of CAMF, a variant of this method that completely neglects spatial interaction, called local CAMF, was proposed. For the tests performed, local CAMF produced almost identical identical results with respect to those obtained with the original version of CAMF. This observation may suggest that the role that spatial interaction plays on the location of optimal sites for afforestation is less relevant than it was initially thought. On the other hand, such similarity of results might be explained by the relationship between the number of cells to be afforested (solution size) and the number of cells in the rasterized representation of the catchment (size of input dataset). In particular, when a small solution size is required for selecting cells in large input datasets, it is likely that the selected cells are located in very disperse areas of the catchment, which decreases the probability of spatial interaction among them. To determine the real relevance of spatial interaction in CAMF, more tests involving larger solution size / input dataset size ratios are necessary. This type of tests proved infeasible for this dissertation given the very long execution times required by the original CAMF in such cases.

The decision criterion considered in CAMF is the amount of sediment delivered to the outlet of a river catchment. While the relevance of such criterion can hardly be argued, there are other objectives that are equally or even more valid in the field of environmental sciences. For instance, given that the primary concern regarding sediment issues is that detached soil remains on the land within the catchment, the minimization of the amount of sediment delivered to the river is an interesting alternative as decision criterion. An additional advantage of considering this criterion instead of sediment delivery to the outlet is that the simulation of sediment flow is stopped whenever it reaches the river system. Besides shortening the execution time required by CAMF, this fact would eliminate the necessity of simulating sediment transportation through the river, avoiding the risk of simulation errors due to uncertainties regarding sediment processes within the river. On the disadvantages side, this criterion shift may require the design and implementation of more complex data structures into the algorithm. Perhaps even more important than assessing sediment delivery to the outlet or to the river is the evaluation of the amount of sediment that flows within the catchment, since the minimization of the latter is one of the ultimate goals of soil conservation. In the ideal, though maybe non-realistic, scenario soil particles do not move from the spatial unit in which they are located. From an environmental point of view, the aspect that really needs to be controlled, i.e., minimized, is the amount of sediment that moves from one spatial unit to another.

7.2.2 Site location for afforestation to optimize multiple on-site criteria

Chapter 4 presents an exploration of several frequently used MCDM methods to rank a set of land units according to their suitability for pine forest with a view to optimize a number of on-site criteria. Since it can be claimed, intuitively, that the ranking produced by all explored methods depends on the values that their parameters take, it would be informative to undertake a sensitivity analysis with regard to the parameter values for each of these MCDM methods. This sensitivity analysis would consist in fixing the values for all except one parameters, while varying the value of the remaining parameter within a sensibly predefined range. Such a process would be repeated for every parameter involved in each method. This technique would allow to formally determine the real impact that tuning the parameters in one or another way has on the output of each method. It is clear as well that sensitivity analysis would ultimately result in more appropriate values for the required parameters.

7.2.3 Land use and trajectory configurations to optimize multiple on-site criteria

In Chapter 5 a Composite Programming model that allows to determine the land use configurations that optimize the multi-dimensional land performance at a regional scale was formulated. In Chapter 6, the same model was applied to search for optimal land use trajectory configurations for a given time span. In both cases Composite Programming was applied to formulate IP models. In particular, its decision variables were restricted to take either 0 or 1 values. This constraint expresses the requirement for a spatial unit to be fully covered by a single LUT or managed according to a single land use trajectory. However formulating the Composite Programming model as a LP model, in which its decision variables can be assigned fractional values, is clearly feasible. In this case, the appropriate constraints would be modified in such a way that the decision variables are allowed to take values in the range [0, 1]. In a given solution to this modified model, the fact that a particular variable takes a fractional value would mean that the corresponding fraction of the spatial unit at stake should be managed according to a certain trajectory. This LP variant of the Composite Programming model would then allow for several trajectories to be assigned to different fractions of the same land unit. This would probably result in a more realistic output , in which land units are not necessarily devoted to a single trajectory.

An additional aspect of the Composite Programming model that is worth to be analyzed is the relationship between regional and local performance. This analysis would involve checking whether a well performing solution at a regional scale corresponds to the case in which most land units present a uniform, good performance or, on the other hand, only a limited group of land units contribute the most to the performance of the region as a whole. In the latter case, emphasis of the land use planner could be on such 'largely contributing' land units, in order to prioritize land use planning on them to take full advantage of their potential while investing limited resources in an efficient way.

The way in which Mathematical Programming models are formulated provides for flexibility that in turn facilitates any necessary adaptations. In this regard, the possibilities to accommodate performance thresholds into the Composite Programming model were analyzed and demonstrated. In the same line of thought, other aspects can be considered in more elaborated variants of this model. For instance, the requirement for a given LUT to cover at least a minimum area threshold can be incorporated by means of adding the appropriate constraints.

Regarding temporal considerations in the Composite Programming model, a seemingly promising approach is to incorporate time directly into the decision variables. This adaptation would shift the meaning of decision variables to express whether a certain LUT should be implemented on a land unit at a given year, which would remove the artificial assumption that land use changes can only occur at fixed points in time. One challenge in this approach would be that the input data requirements might become very demanding, especially in cases in which a fine grained temporal resolution is a requisite.

Besides these initial steps, from a more general point of view, the full integration of the temporal dimension into methods to support afforestation planning would allow them to tackle more challenging questions, such as "When (in which specific year) should a forest be established?". Providing answers to such questions would require to consider additional and perhaps more complex aspects, like uncertainties regarding future climate and the influence of gradual tree growth on land performance. Additional time related questions worth to be studied

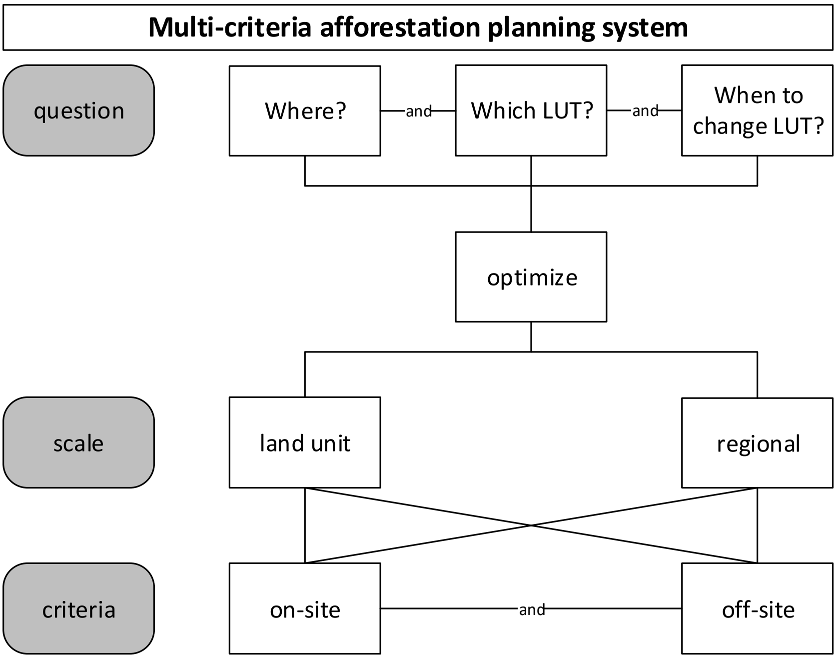

Figure 7.2: Schematic view of a comprehensive afforestation planning framework.

would be “When should a forest be harvested?”, always with a view to optimize predefined criteria.

7.2.4 Possibilities and challenges towards a comprehensive and flexible afforestation planning framework

To summarize the discussion in this section, Figure 7.2 shows a schematic view of the components (and the inter-relationships among them) of a comprehensive afforestation planning system according to the perspectives discussed in this dissertation.

The afforestation planning framework in Figure 7.2 is organized in three levels: the question level, the scale level and the criteria level, as shown by the shaded boxes to the left. In the first level from top to bottom, this framework should be able to deal with general questions that simultaneously consider several aspects: where (site location) and with which tree species to afforest?, whether or when the should the forest be harvested? should the previous tree species be replanted or should a new species be used? All these questions have been, in one way or another, dealt with in this dissertation although each of them has been addressed separately. To consider them in a simultaneous way would require an ensemble of diverse methods like the ones presented individually in this work. Special care would be necessary to ensure a coordinated functioning of such an ensemble with emphasis on establishing the appropriate communication channels among the different techniques involved.

The answers to the questions stated above are oriented to optimization either on a 'per land unit' basis or at a regional scale. Regional optimization may seem more relevant at first sight, specially for strategic afforestation planning in, e.g., governmental organizations. However, regional optimization is not so obvious when dealing with site location problems, in which the performance of each land unit is assessed separately to determine whether it is selected for afforestation or not. In such cases, optimization at land unit scale is intuitively more natural.

Probably the most obvious requirement for a comprehensive afforestation planning system is the ability to consider more than one criterion. The requirements of the users of this system may indicate that the set of considered criteria corresponds to a mixture of both on- and off-site attributes. One of the differences between these two types of criteria is the availability of the attribute information. When dealing with on-site criteria, typically the information about performance levels can be computed in advance, in such a way that it is available before the execution of the corresponding method starts. Regarding off-site criteria, such information cannot be computed beforehand, since the performance levels of any land unit depend on changes that may occur in other units. Therefore, off-site performance information typically is computed during the execution of the method, like in the case of CAMF and its variants. It is also typical that the computation of performance values requires the implementation of an algorithmic procedure. This aspect can become a limiting factor in case off-site criteria are to be incorporated in Mathematical Programming models, which do not follow an algorithmic paradigm, although the incorporation of an off-site attribute in a single criterion Mathematical Programming model has been proved feasible by Vanegas et al. (2012). In this dissertation on-site and off-site criteria were considered as separate problems and tackled using different methods.

The work conducted in this dissertation can be seen as the initial steps towards a comprehensive decision support framework for afforestation planning. The implementation of such framework would require an ensemble of the studied methods, with the suggested extensions, plus other additional existing or to be developed techniques aiming at related purposes. The envisaged integral framework would cover as many as possible of the relevant aspects involved in afforestation planning. In summary, it would be aimed at answering questions like: In which specific year, with which species and where a forest should be established and for how long should it be kept in place, in order to simultaneously optimize several conflicting on- and off-site criteria, taking into account performance and area restrictions.

Bibliography

Aerts, J., Van Herwijnen, M., Janssen, R., and Stewart, T. (2005). Evaluating spatial design techniques for solving land-use allocation problems. *Journal of Environmental Planning and Management*, 48(1):121–142.

Ananda, J. and Herath, G. (2009). A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological Economics*, 68(10):2535–2548.

André, F. and Romero, C. (2008). Computing compromise solutions: on the connections between compromise programming and composite programming. *Applied Mathematics and Computation*, 195(1):1–10.

Bardossy, A. and Bogardi, I. (1983). Network design for the spatial estimation of environmental variables. *Applied Mathematics and Computation*, 12(4):339–365.

Belton, V. and Stewart, T. (2002). *Multiple criteria decision analysis: an integrated approach*. Springer Science & Business Media.

Blasco, F., Cuchillo-Ibáñez, E., Morón, M., and Romero, C. (1999). On the monotonicity of the compromise set in multicriteria problems. *Journal of Optimization Theory and Applications*, 102(1):69–82.

Brans, J. and Mareschal, B. (2005). Promethee methods. In *Multiple Criteria Decision Analysis: State of the Art Surveys*, volume 78 of *International Series in Operations Research & Management Science*, pages 163–186. Springer New York.

Brans, J. and Vincke, P. (1985). A preference ranking organisation method: The promethee method for multiple criteria decision-making. *Management science*, 31(6):647–656.

Brown, D. (1994). Predicting vegetation types at treeline using topography and biophysical disturbance variables. *Journal of Vegetation Science*, pages 641–656.

Buytaert, W., Célleri, R., De Bièvre, B., Cisneros, F., Wyseure, G., Deckers, J., and Hofstede, R. (2006). Human impact on the hydrology of the andean páramos. *Earth-Science Reviews*, 79(1):53–72.

Cao, K., Huang, B., Wang, S., and Lin, H. (2012). Sustainable land use optimization using boundary-based fast genetic algorithm. *Computers, Environment and Urban Systems*, 36(3):257–269.

Chang, N., Wen, C., and Wu, S. (1995). Optimal management of environmental and land resources in a reservoir watershed by multiobjective programming. *Journal of Environmental Management*, 44(2):144–161.

Chang, Y.-C. and Ko, T.-T. (2014). An interactive dynamic multi-objective programming model to support better land use planning. *Land Use Policy*, 36:13–22.

Charnes, A. and Collomb, B. (1972). Optimal economic stabilization policy: Linear goal-interval programming models. *Socio-Economic Planning Sciences*, 6(4):431–435.

Charnes, A. and Cooper, W. (1957). Management models and industrial applications of linear programming. *Management Science*, 4(1):38–91.

Charnes, A., Cooper, W., and Ferguson, R. (1955). Optimal estimation of executive compensation by linear programming. *Management Science*, 1(2):138–151.

Chen, J., Zhang, X., and Zhu, Q. (2011). Multi-objective decision making for land use planning with ordered weighted averaging method. *Systems Engineering Procedia*, 2:434–440.

Chen, Y., Yu, J., and Khan, S. (2010). Spatial sensitivity analysis of multi-criteria weights in gis-based land suitability evaluation. *Environmental Modelling & Software*, 25(12):1582–1591.

Cisneros, P. (1999). Mapping of soil erosion risk zones in a part of the watershed of rio paute (ecuador).

Costa-Cabral, M. and Burges, S. (1994). Digital elevation model networks (demon): A model of flow over hillslopes for computation of contributing and dispersal areas. *Water Resources Research*, 30(6):1681–1692.

Costin, A. (1980). Runoff and soil and nutrient losses from an improved pasture at ginninderra, southern tablelands, new south wales. *Crop and Pasture Science*, 31(3):533–546.

Darradi, Y., Saur, E., Laplana, R., Lescot, J.-M., Kuentz, V., and Meyer, B. (2012). Optimizing the environmental performance of agricultural activities: A case study in la boulouze watershed. *Ecological Indicators*, 22:27–37.

De Keukelaere, L. (2013). Modeling of runoff and sediment production to optimize land use distribution in the southern andes of ecuador.

De Meyer, A., Estrella, R., Jacxsens, P., Deckers, J., Van Rompaey, A., and Van Orshoven, J. (2013). A conceptual framework and its software implementation to generate spatial decision support systems for land use planning. *Land Use Policy*, 35:271–282.

Diaz-Balteiro, L. and Romero, C. (2008). Making forestry decisions with multiple criteria: a review and an assessment. *Forest Ecology and Management*, 255(8):3222–3241.

Doumpos, M. and Zopounidis, C. (2002). *Multicriteria decision aid classification methods*, volume 73. Springer Science & Business Media.

Drzewiecki, W. and Mularz, S. (2008). Simulation of water soil erosion effects on sediment delivery to dobczyce reservoir. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37:B8.

Duh, J.-D. and Brown, D. (2007). Knowledge-informed pareto simulated annealing for multi-objective spatial allocation. *Computers, Environment and Urban Systems*, 31(3):253–281.

Eliasson, Å., Jones, R., Nachtergaele, F., Rossiter, D., Terres, J., Van Orshoven, J., Van Velthuizen, H., Böttcher, K., Haastrup, P., and Le Bas, C. (2010). Common criteria for the redefinition of intermediate less favoured areas in the european union. *Environmental Science & Policy*, 13(8):766–777.

Estrella, R., Cattrysse, D., and Van Orshoven, J. (2014a). Comparison of three ideal point-based multi-criteria decision methods for afforestation planning. *Forests*, 5(12):3222–3240.

Estrella, R., Delabastita, W., Wijffels, A., Cattrysse, D., and Van Orshoven, J. (2014b). Comparison of multicriteria decision making methods for selection of afforestation sites. *Revue Internationale de Géomatique*, 24(2):143–157.

Estrella, R., Vanegas, P., Cattrysse, D., and Van Orshoven, J. (2014c). Trading off accuracy and computational efficiency of an afforestation site location method for minimizing sediment yield in a river catchment. In *Proceedings of GEOProcessing 2014: The Sixth International Conference on Advanced Geographic Information Systems, Applications, and Services*, pages 94–100.

FAO (1976). *A framework for land evaluation (Soils Bulletin 32).* Food and Agricultural Organization of the United Nations.

FAO (1980). *A provisional methodology for soil degradation assessment.* Food and Agriculture Organization of the United Nations.

FAO (1993). *Forest resources assessment 1990 - Tropical countries (FAO Forestry Paper 112).* Food and Agricultural Organization of the United Nations.

FAO (2007). *Land evaluation - towards a revised framework (Land and Water Discussion Paper 6).* Food and Agricultural Organization of the United Nations.

Fischer, D. and Church, R. (2003). Clustering and compactness in reserve site selection: an extension of the biodiversity management area selection model. *Forest Science*, 49(4):555–565.

Freimer, M. and Yu, P. (1976). Some new results on compromise solutions for group decision problems. *Management Science*, 22(6):688–693.

Fu, B. and Gulinck, H. (1994). Land evaluation in an area of severe erosion: the loess plateau of china. *Land Degradation & Development*, 5(1):33–40.

Gal, T., Stewart, T., and Hanne, T. (2013). *Multicriteria decision making: advances in MCDM models, algorithms, theory, and applications*, volume 21. Springer Science & Business Media.

Gao, W. and Lesht, B. (1997). Model inversion of satellite-measured reflectances for obtaining surface biophysical and bidirectional reflectance characteristics of grassland. *Remote Sensing of Environment*, 59(3):461–471.

Gibney, R. and Shang, J. (2007). Decision making in academia: A case of the dean selection process. *Mathematical and Computer Modelling*, 46(7):1030–1040.

Gilliams, S., Raymaekers, D., Muys, B., and Van Orshoven, J. (2005a). Comparing multiple criteria decision methods to extend a geographical information system on afforestation. *Computers and Electronics in Agriculture*, 49(1):142–158.

Gilliams, S., Van Orshoven, J., Muys, B., Kros, H., Heil, W., and Van Deursen, W. (2005b). Afforest sdss: a metamodel based spatial decision support system for afforestation of agricultural land. *New Forests*, 30(1):33–53.

Godschalk, D. (2004). Land use planning challenges: Coping with conflicts in visions of sustainable development and livable communities. *Journal of the American Planning Association*, 70(1):5–13.

Heil, G., Muys, B., and Hansen, K. (2007). *Environmental effects of afforestation in north-western Europe: From field observations to decision support*, volume 1. Springer Science & Business Media.

Hewawasam, T., von Blanckenburg, F., Schaller, M., and Kubik, P. (2003). Increase of human over natural erosion rates in tropical highlands constrained by cosmogenic nuclides. *Geology*, 31(7):597–600.

Ho, W., Xu, X., and Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research*, 202(1):16–24.

Hobbs, B. and Meier, P. (1994). Multicriteria methods for resource planning: an experimental comparison. *IEEE Transactions on Power Systems*, 9(4):1811–1817.

Holzkämper, A. and Seppelt, R. (2007). A generic tool for optimising land-use patterns and landscape structures. *Environmental Modelling & Software*, 22(12):1801–1804.

Joerin, F., Thériault, M., and Musy, A. (2001). Using gis and outranking multicriteria analysis for land-use suitability assessment. *International Journal of Geographical Information Science*, 15(2):153–174.

Kahn, A. (1962). Topological sorting of large networks. *Communications of the ACM*, 5(11):558–562.

Kangas, J. and Kangas, A. (2005). Multiple criteria decision support in forest management: the approach, methods applied, and experiences gained. *Forest Ecology and Management*, 207(1):133–143.

Keeney, R. L. and Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge University Press.

Kiker, G. A., Bridges, T. S., Varghese, A., Seager, T. P., and Linkov, I. (2005). Application of multicriteria decision analysis in environmental decision making. *Integrated Environmental Assessment and Management*, 1(2):95–108.

Krcmar, E., van Kooten, G. C., and Vertinsky, I. (2005). Managing forest and marginal agricultural land for multiple tradeoffs: compromising on economic, carbon and structural diversity objectives. *Ecological Modelling*, 185(2):451–468.

Lahdelma, R. and Salminen, P. (2001). Smaa-2: Stochastic multi-criteria acceptability analysis for group decision making. *Operations Research*, 49(3):444–454.

Lahdelma, R., Salminen, P., and Hokkanen, J. (2000). Using multicriteria methods in environmental planning and management. *Environmental management*, 26(6):595–605.

Malczewski, J. (1999). *GIS and multicriteria decision analysis*. John Wiley & Sons.

MEA (2005). Linking ecosystem services and human well-being. Washington DC.

Mendoza, G. and Martins, H. (2006). Multi-criteria decision analysis in natural resource management: a critical review of methods and new modelling paradigms. *Forest Ecology and Management*, 230(1):1–22.

Messac, A., Ismail-Yahaya, A., and Mattson, C. A. (2003). The normalized normal constraint method for generating the pareto frontier. *Structural and Multidisciplinary Optimization*, 25(2):86–98.

Molina, A., Govers, G., Poesen, J., Van Hemelryck, H., De Bièvre, B., and Vanacker, V. (2008). Environmental factors controlling spatial variation in sediment yield in a central andean mountain area. *Geomorphology*, 98(3):176–186.

Morgan, R. (2009). *Soil erosion and conservation*. John Wiley & Sons.

Nearing, M., Jetten, V., Baffaut, C., Cerdan, O., Couturier, A., Hernandez, M., Le Bissonnais, Y., Nichols, M., Nunes, J., and Renschler, C. (2005). Modeling response of soil erosion and runoff to changes in precipitation and cover. *Catena*, 61(2):131–154.

O'Callaghan, J. and Mark, D. (1984). The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, 28(3):323–344.

Opricovic, S. and Tzeng, G.-H. (2004). Compromise solution by mcdm methods: A comparative analysis of vikor and topsis. *European Journal of Operational Research*, 156(2):445–455.

Palmieri, A., Shah, F., and Dinar, A. (2001). Economics of reservoir sedimentation and sustainable management of dams. *Journal of Environmental Management*, 61(2):149–163.

Piégay, H., Walling, D., Landon, N., He, Q., Liébault, F., and Petiot, R. (2004). Contemporary changes in sediment yield in an alpine mountain basin due to afforestation (the upper drôme in france). *Catena*, 55(2):183–212.

Pimentel, D., Harvey, C., Resosudarmo, P., Sinclair, K., Kurz, D., McNair, M., Crist, S., Shpritz, L., Fitton, L., Saffouri, R., and Blair, R. (1995). Environmental and economic costs of soil erosion and conservation benefits. *Science*, 267(5201):1117–1123.

Pohekar, S. and Ramachandran, M. (2004). Application of multi-criteria decision making to sustainable energy planning — a review. *Renewable and Sustainable Energy Reviews*, 8(4):365–381.

Postma, G., Kleinhans, M., Meijer, P., and Eggenhuisen, J. (2008). Sediment transport in analogue flume models compared with real-world sedimentary systems: A new look at scaling evolution of sedimentary systems in a flume. *Sedimentology*, 55(6):1541–1557.

Privette, J., Eck, T., and Deering, D. (1997). Estimating spectral albedo and nadir reflectance through inversion of simple brdf models with avhrr/modis-like data. *Journal of Geophysical Research*, 102(D24).

PROMAS (2005). Plan de manejo integral de la microcuenca del rio tabacay (in spanish).

Quinn, P., Beven, K., Chevallier, P., and Planchon, O. (1991). Prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, 5(1):59–79.

Ramanathan, R. (2001). A note on the use of the analytic hierarchy process for environmental impact assessment. *Journal of Environmental Management*, 63(1):27–35.

Raymaekers, D. (2003). Meervoudige doelplanning bij de bebossing van landbouwgronden: naar een ruimtelijk beslissingsondersteunend systeem.

Renard, K., Foster, G., Weesies, G., and Porter, J. (1991). Rusle: Revised universal soil loss equation. *Journal of Soil and Water Conservation*, 46(1):30–33.

Roering, J., Kirchner, J., and Dietrich, W. (1999). Evidence for nonlinear, diffusive sediment transport on hillslopes and implications for landscape morphology. *Water Resources Research*, 35(3):853–870.

Romero, C. and Rehman, T. (2003). *Multiple criteria analysis for agricultural decisions*. Elsevier.

Rossmiller, R. (1982). Rational formula revisited. In DeGroot, W., editor, *Stormwater Detention Facilities: Planning, Design, Operation, and Maintenance*, pages 146–162.

Roy, B. (1991). The outranking approach and the foundations of electre methods. *Theory and decision*, 31(1):49–73.

Saaty, T. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281.

Salminen, P., Hokkanen, J., and Lahdelma, R. (1998). Comparing multicriteria methods in the context of environmental problems. *European Journal of Operational Research*, 104(3):485–496.

Schädler, S., Finkel, M., Bleicher, A., Morio, M., and Gross, M. (2013). Spatially explicit computation of sustainability indicator values for the automated assessment of land-use options. *Landscape and Urban Planning*, 111:34–45.

Schädler, S., Morio, M., Bartke, S., and Finkel, M. (2012). Integrated planning and spatial evaluation of megasite remediation and reuse options. *Journal of Contaminant Hydrology*, 127(1):88–100.

Tarboton, D. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, 33(2):309–319.

Tenerelli, P. and Carver, S. (2012). Multi-criteria, multi-objective and uncertainty analysis for agro-energy spatial modelling. *Applied Geography*, 32(2):724–736.

Tervonen, T. (2014). Jsmaa: open source software for smaa computations. *International Journal of Systems Science*, 45(1):69–81.

Van Oost, K., Govers, G., and Desmet, P. (2000). Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. *Landscape Ecology*, 15(6):577–589.

Van Rompaey, A., Bazzoffi, P., Jones, R., and Montanarella, L. (2005). Modeling sediment yields in italian catchments. *Geomorphology*, 65(1):157–169.

Van Rompaey, A., Krasa, J., Dostal, T., and Govers, G. (2003). Modelling sediment supply to rivers and reservoirs in eastern europe during and after the collectivisation period. *Hydrobiologia*, 494(1-3):169–176.

Van Rompaey, A., Verstraeten, G., Van Oost, K., Govers, G., and Poesen, J. (2001). Modelling mean annual sediment yield using a distributed approach. *Earth Surface Processes and Landforms*, 26(11):1221–1236.

Vanacker, V., Molina, A., Govers, G., Poesen, J., and Deckers, J. (2007). Spatial variation of suspended sediment concentrations in a tropical andean river system: The paute river, southern ecuador. *Geomorphology*, 87(1):53–67.

Vanegas, P. (2010). A spatially explicit approach to the site location problem in raster maps with application to afforestation.

Vanegas, P., Cattrysse, D., and Van Orshoven, J. (2008). Comparing exact and heuristic methods for site location based on multiple attributes: An afforestation application. In *Computational Science and Its Applications–ICCSA 2008*, pages 389–404. Springer.

Vanegas, P., Cattrysse, D., and Van Orshoven, J. (2012). Allocating reforestation areas for sediment flow minimization: an integer programming formulation and a heuristic solution method. *Optimization and Engineering*, 13(2):247–269.

Verstegen, J. A., Karssenberg, D., van der Hilst, F., and Faaij, A. (2012). Spatio-temporal uncertainty in spatial decision support systems: A case study of changing land availability for bioenergy crops in mozambique. *Computers, Environment and Urban Systems*, 36(1):30–42.

Verstraeten, G., Van Oost, K., Van Rompaey, A., Poesen, J., and Govers, G. (2002). Evaluating an integrated approach to catchment management to reduce soil loss and sediment pollution through modelling. *Soil Use and Management*, 18(4):386–394.

Vought, L., Pinay, G., Fuglsang, A., and Ruffinoni, C. (1995). Structure and function of buffer strips from a water quality perspective in agricultural landscapes. *Landscape and Urban Planning*, 31(1):323–331.

Wang, J., Jing, Y., Zhang, C., and Zhao, J. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*, 13(9):2263–2278.

Weiss, M., Baret, F., Myneni, R., Pragnère, A., and Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20(1):3–22.

Wijffels, A. and Van Orshoven, J. (2009). Contribution of afforestation to the enhancement of physical and socio-economic land performance in the southern andes of ecuador: Assessment, modelling and planning support (unpublished).

Williams, J. (2002). A zero-one programming model for contiguous land acquisition. *Geographical Analysis*, 34(4):330–349.

Wilson, J. and Gallant, J. (2000). *Terrain analysis: principles and applications.* Wiley.

Winston, W. and Goldberg, J. (2004). *Operations Research: applications and algorithms*, volume 3. Duxbury Press Boston.

Wischmeier, W. and Smith, D. (1978). Predicting rainfall erosion losses: a guide to conservation planning. In of Agriculture, U. D., editor, *Agriculture Handbook*.

Xiao, N. (2008). A unified conceptual framework for geographical optimization using evolutionary algorithms. *Annals of the Association of American Geographers*, 98(4):795–817.

Yu, P. (1973). A class of solutions for group decision problems. *Management Science*, 19(8):936–946.

Yu, X., Zhang, X., Li, J., Zhang, M., and Xie, Y. (2006). Effects of vegetation cover and precipitation on the process of sediment produced by erosion in a small watershed of loess region. *Acta Ecologica Sinica*, 26(1):1–8.

Zanakis, S. H., Solomon, A., Wishart, N., and Dublish, S. (1998). Multi-attribute decision making: A simulation comparison of select methods. *European Journal of Operational Research*, 107(3):507–529.

Zarco-Tejada, P., Miller, J., Noland, T., Mohammed, G., and Sampson, P. (2001). Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1491–1507.

Zeleny, M. (1973). Compromise programming. *Multiple Criteria Decision Making*, 286.

Zopounidis, C. and Doumpos, M. (2002). Multicriteria classification and sorting methods: a literature review. *European Journal of Operational Research*, 138(2):229–246.

List of publications

Articles in internationally reviewed academic journals

Estrella, R., Delabastita, W., Wijffels, A., Cattrysse, D., Van Orshoven, J. (2014). Comparison of multicriteria decision making methods for selection of afforestation sites. *Revue Internationale de Géomatique*, 24 (2), 143-157.

Estrella, R., Cattrysse, D., Van Orshoven, J. (2014). Comparison of three ideal point-based multi-criteria decision methods for afforestation planning. *Forests*, 5 (12), 3222-3240.

Van Orshoven, J., Kint, V., Wijffels, A., Estrella, R., Bencsik, G., Vanegas, P., Muys, B., Cattrysse, D., Dondeyne, S. (2011). Upgrading geographic information systems to spatial decision support systems. *Mathematical and Computational Forestry & Natural-Resource Sciences*, 3 (1), 36-41.

De Meyer, A., Estrella, R., Jacxsens, P., Deckers, S., Van Rompaey, A., Van Orshoven, J. (2013). A conceptual framework and its software implementation to generate spatial decision support systems for land use planning. *Land Use Policy*, 35, 271-282.

Van Orshoven, J., Hubeau, M., De Meyer, A., Estrella, R., Steenberghen, T. (2014). Student satisfaction and performance in relation to software-independent tutorials of GI S&T. *Revue Internationale de Géomatique*, 24 (2), 251-268.

Articles submitted to internationally reviewed academic journas, under review

Estrella, R., Cattrysse, D., Van Orshoven, J. (under review). A Composite Programming model to determine land use trajectories for optimizing regionally

integrated ecosystem services delivery. *Forests: Special Issue Ecosystem Services from Forests.*

Articles at international scientific conferences, published in full in proceedings

Estrella, R., Delabastita, W., Wijffels, A., Van Orshoven, J. (2012). Comparison of discrete multi-criteria decision making methods for selection of afforestation sites. *Actes de la Conférence internationale de Géomatique et Analyse Spatiale 2012.* Liege, Belgium, 6-9 November 2012 (pp. 182-197). Unité de Géomatique, Département de géographie, Université de Liege.

Estrella, R., Vanegas, P., Cattrysse, D., Van Orshoven, J. (2014). Trading off accuracy and computational efficiency of an afforestation site location method for minimizing sediment yield in a river catchment. In Rückemann, C. (Ed.), Doytsher, Y. (Ed.), *Proceedings of GEOProcessing 2014: The Sixth International Conference on Advanced Geographic Information Systems, Applications, and Services.* Barcelona, Spain, 23-27 March 2014 (pp. 94-100). International Academy, Research, and Industry Association (IARIA).

Van Orshoven, J., Kint, V., Wijffels, A., Estrella, R., Bencsik, G., Vanegas, P., Muys, B., Cattrysse, D., Dondeyne, S. (2010). Upgrading geographic information systems to spatio-temporal decision support systems. In Falcao, A. (Ed.), Rosset, C. (Ed.), *Proceedings of the workshop on Decision Support Systems in Sustainable Forest Management (DSFM 2010).* Lisbon, 19-21 April 2010 (article number OP 7).

Van Orshoven, J., Hubeau, M., Estrella, R., De Meyer, A., Vileikis, O., Steenberghen, T. (2011). In-class or self-study of GI S&T ? Five years of experience with FOSS4G in teaching GI S&T. *Proceedings of the LeGIO-workshop. GIS-education in a changing academic environment.* Leuven, Belgium, 18 November 2011 (pp. 97-107). SADL - Spatial Applications Division Leuven.

Abstracts presented at international scientific conferences

Van Orshoven, J., Wijffels, A., Estrella, R., Cattrysse, D., Cisneros, F., Muys, B. (2009). From land evaluation to spatio-temporal decision support on afforestation. *Proceedings of the AGILE 2009 conference.* Hannover, Germany, 1-5 June, 2009.

Articles in preparation

Zhang, K., Estrella, R., Cattrysse, D., Van Orshoven, J. Determining the optimal afforestation pattern in a mountainous catchment with a multiple flow direction heuristic method.